

Reactive Attitudes e Grammatica dell'Agentività

I. Il problema della presupposizione grammaticale nella filosofia analitica

1.1 Una forma di cecità non accidentale

C'è una categoria di errori filosofici che non si commettono per mancanza di acutezza o per carenze nell'analisi: si commettono perché la struttura stessa del medium in cui il pensiero si svolge rende certi problemi invisibili. Non si tratta dell'errore ordinario — la premessa falsa, il ragionamento invalido, la confusione concettuale — ma di qualcosa di più radicale: la presenza, nella struttura grammaticale della lingua in cui si filosofa, di presupposizioni ontologiche che non vengono mai tematizzate perché non vengono mai percepite come scelte. Esse operano al di sotto della soglia della deliberazione concettuale, nel livello in cui la lingua funziona come strumento trasparente piuttosto che come oggetto di analisi. Chiamare queste presupposizioni «invisibili» non è una metafora: esse sono invisibili nel senso preciso in cui sono invisibili le strutture cognitive che rendono possibile la visione — non sono viste, ma tutto ciò che è visto dipende da esse.

Il presente testo si occupa di una di queste presupposizioni, che è al tempo stesso tra le più pervasive e le meno esplorate nella tradizione analitica: la presupposizione che l'agentività sia una proprietà binaria, che una certa entità o possiede o non possiede, in un senso univoco e non ulteriormente analizzabile. Questa presupposizione non è formulata esplicitamente nella letteratura analitica sull'azione, sulla responsabilità e sul libero arbitrio: essa non appare come premessa di nessun argomento, non viene difesa contro obiezioni, non riceve motivazione. Essa opera piuttosto come sfondo strutturale — come quella che Wittgenstein chiamava «la forma della nostra grammatica» — rispetto al quale tutti gli argomenti espliciti vengono formulati e valutati.

La tesi di questo testo è che questa presupposizione non riflette la struttura dell'agentività umana, ma la struttura grammaticale delle lingue indoeuropee, che codificano la partecipazione di un'entità a un evento attraverso la categoria del soggetto nominativo — una categoria che si applica in modo identico all'originatore causale intenzionale di un'azione deliberata, all'esecutore di un'azione non pienamente volontaria, all'agente ipnotico che «fa» qualcosa senza vera intenzionalità, al soggetto che sperimenta uno stato psicologico senza averlo prodotto. In tutte queste situazioni profondamente diverse, il soggetto grammaticale è il soggetto: la struttura morfologica dell'enunciazione non distingue, non analizza, non differenzia. E poiché la tradizione analitica ha pensato in greco, latino, tedesco, inglese, francese — tutte lingue che condividono questa struttura fondamentale — essa ha ereditato la presupposizione senza mai interrogarla.

1.2 Perché la tradizione analitica è il caso più istruttivo

Vi sono buone ragioni per concentrare l'analisi sulla tradizione analitica piuttosto che sulla filosofia occidentale in generale. La prima è che la filosofia analitica ha sviluppato la propria metodologia con un'attenzione all'analisi del linguaggio ordinario e alla chiarificazione concettuale che non ha precedenti nella storia della filosofia: essa ha esplicitamente tematizzato il rischio delle presupposizioni nascoste nel linguaggio e ha prodotto strumenti per identificarle e dissolverle. Eppure — e qui si trova il paradosso istruttivo — questa tradizione ha analizzato con cura le presupposizioni del linguaggio ordinario senza mai analizzare le presupposizioni grammaticali del linguaggio in cui essa stessa si svolgeva. Il programma russelliano di analisi logica, il progetto wittgensteiniano di terapia linguistica, il programma oxoniense dell'*ordinary language philosophy*: nessuno di questi ha mai considerato che il linguaggio ordinario di cui si occupavano fosse il linguaggio ordinario di parlanti di lingue indoeuropee, e che le sue strutture grammaticali portassero presupposizioni ontologiche specifiche di una famiglia linguistica, non strutture universali della razionalità o dell'esperienza umana.

La seconda ragione è che la tradizione analitica ha prodotto i contributi più sofisticati sul problema specifico che qui si esamina — la struttura della responsabilità, degli atteggiamenti reattivi, del libero arbitrio — e questi contributi sono formulati in un linguaggio tecnico abbastanza preciso da permettere di

identificare con chiarezza dove le presupposizioni grammaticali operano. Quando Strawson scrive che gli atteggiamenti reattivi sono risposte alla «qualità della volontà» dell'agente, o quando Davidson scrive che un'azione è un evento causato «nel modo giusto» dagli stati mentali dell'agente, o quando Frankfurt distingue tra desideri di primo e secondo ordine, si ha una terminologia abbastanza precisa da permettere di vedere dove la struttura grammaticale ha pre-selezionato le categorie disponibili. Le formulazioni vaghe o metaforiche nascondono le presupposizioni grammaticali con troppa efficacia: le formulazioni analitiche precise le espongono.

1.3 Il problema metodologico: dalla linguistica alla filosofia

Prima di procedere nell'analisi sostantiva, è necessario affrontare una questione metodologica che riguarda il tipo di argomento che si intende sviluppare e il suo statuto epistemologico. Ci sono almeno tre modi distinti in cui la struttura grammaticale di una lingua può essere rilevante per la filosofia, e confonderli — come avviene con notevole frequenza nella letteratura — produce argomenti che sembrano forti ma che si rivelano vulnerabili ad obiezioni ovvie.

Il primo modo è causale-cognitivo: la struttura grammaticale di una lingua produce, nei suoi parlanti, effetti cognitivi misurabili che determinano quali concetti sono più accessibili, quali inferenze sono più automatiche, quali distinzioni sono più facili o difficili. Questo è il dominio della linguistica cognitiva e della psicologia cross-linguistica, e le affermazioni in questo dominio hanno statuto di ipotesi empiriche verificabili sperimentalmente. La letteratura in questo campo — da Whorf a Slobin a Levinson a Boroditsky — ha prodotto risultati variabili e controversi. Alcuni effetti sono documentati con ragionevole robustezza; altri non si replicano; il meccanismo causale rimane in molti casi oscuro. Chi voglia sostenere la tesi causale-cognitiva deve fare i conti con questa letteratura in tutta la sua complessità, e non può semplicemente invocarla come supporto senza specificare quali risultati specifici siano rilevanti per quale affermazione specifica.

Il secondo modo è storico-ricostruttivo: indipendentemente dagli effetti cognitivi, la struttura grammaticale delle lingue in cui certi problemi filosofici sono stati formulati per la prima volta ha co-determinato la forma in cui quei problemi sono stati posti, rendendone alcune formulazioni più naturali e altre

più faticose, e questo si può mostrare attraverso analisi storico-testuali. Questo è un tipo di argomento che ha statuto storico-interpretativo, non empirico: la sua forza dipende dalla capacità di mostrare, nei testi effettivi della tradizione, tracce riconoscibili dell'influenza grammaticale. Non richiede che si dimostri alcun effetto cognitivo generale: richiede soltanto che si mostri come certe formulazioni abbiano reso certi problemi più visibili e altri meno.

Il terzo modo è filosofico-concettuale: indipendentemente dagli effetti cognitivi e dalla storia effettiva della tradizione, la comparazione tra grammatiche diverse mostra che certi problemi filosofici possono essere formulati in modi più articolati quando si dispone di categorie grammaticali più ricche, e questo ha conseguenze normative per come quei problemi dovrebbero essere formulati. Questo è il tipo di argomento più robusto dal punto di vista filosofico, perché non richiede di dimostrare alcuna tesi empirica sulla cognizione o sulla storia: richiede soltanto di mostrare che le categorie grammaticali di certe lingue articolano distinzioni che sono indipendentemente motivate da ragioni filosofiche e che il sistema binario delle lingue indoeuropee non articola.

L'argomento di questo testo procede principalmente nel terzo modo, con apporti dal secondo. Si mostra che il sistema morfologico georgiano articola distinzioni dell'agentività che sono indipendentemente motivate da ragioni filosofiche — ragioni che la tradizione analitica ha riconosciuto in modo disperso e mai sistematico — e che la loro mancata articolazione sistematica nella tradizione è spiegabile attraverso il secondo tipo di argomento, quello storico-ricostruttivo. Non si fa affidamento sul primo tipo di argomento — quello causale-cognitivo — come premessa necessaria, anche se alcune ricerche in quel dominio sono compatibili con le conclusioni di questo testo.

Questa scelta metodologica ha un costo: essa non permette di dire che il pensiero dei filosofi analitici è stato «determinato» o «causato» dalla loro grammatica. Ma questo costo è accettabile, anzi preferibile, perché la tesi deterministica è sia empiricamente non dimostrata sia filosoficamente problematica in un modo che indebolirebbe l'intero argomento. Ciò che si intende mostrare è più preciso e più difendibile: che la grammatica binaria dell'agentività ha reso certi problemi più difficili da vedere e certi strumenti concettuali meno disponibili, e che la comparazione con il sistema georgiano

permette di formulare quegli strumenti in modo esplicito e sistematico.

1.4 Merleau-Ponty e l'agentività incorporata: un anticipo filosofico

Prima di procedere all'analisi del sistema georgiano, è utile segnalare che la tradizione filosofica continentale ha prodotto, in modo indipendente dalla tipologia linguistica, intuizioni sull'agentività che convergono con il quadro multidimensionale che questo testo sviluppa. Maurice Merleau-Ponty, nella *Fenomenologia della percezione* (1945), ha elaborato una critica della concezione soggetto-centrica dell'azione che anticipa in molti punti le insufficienze del modello binario indoeuropeo. Per Merleau-Ponty, l'azione non è l'effetto di una volontà che comanda un corpo passivo: è una struttura di risposta in cui il corpo-soggetto si orienta verso il mondo attraverso schemi motori acquisiti che non sono pienamente intenzionali nel senso cartesiano, ma non sono nemmeno puramente meccanici. Il concetto di «intenzionalità motoria» — la direzionalità del corpo verso il suo ambiente senza la mediazione esplicita della rappresentazione cosciente — cattura esattamente quella zona dell'agentività che le lingue indoeuropee confondono nel soggetto nominativo.

La ragione per cui questo anticipo fenomenologico non ha trovato sbocco nella filosofia analitica dell'azione è in parte istituzionale — le due tradizioni hanno comunicato poco — e in parte grammaticale. Merleau-Ponty scrive in francese, una lingua con lo stesso schema nominativo-accusativo del latino e del greco che questa trattazione esamina. Per esprimere l'agentività intermedia che la voce media greca o il dativo georgiano grammaticalizzano, egli deve costruire costruzioni perifrasi complesse e spesso oscure — «il corpo che si muove verso», «l'intenzionalità operante», «l'arco intenzionale» — che cercano di dire ciò che la grammatica francese non dice automaticamente. Il risultato è una prosa di grande potenza filosofica ma di difficile esegesi, la cui oscurità è in parte intrinseca alla difficoltà di esprimere in una grammatica binaria strutture che richiederebbero una grammatica graduale. Il sistema georgiano non avrebbe reso Merleau-Ponty superfluo; avrebbe reso la sua prosa più perspicua.

Questa convergenza tra fenomenologia continentale e tipologia linguistica non è accidentale. Entrambe identificano — attraverso strumenti e metodi molto diversi — la stessa

insufficienza del modello binario dell'agentività: la sua incapacità di rendere conto delle zone intermedie tra l'azione intenzionale piena e la passività totale, zone che costituiscono, come l'esperienza ordinaria conferma, la stragrande maggioranza degli eventi che co-determiniamo. Questa convergenza è anche un argomento indiretto a favore della solidità filosofica del quadro multidimensionale: esso non è un'importazione esotica dalla tipologia linguistica, ma articola sistematicamente una struttura che pensatori diversi, in tradizioni diverse, hanno riconosciuto come necessaria attraverso approcci indipendenti.

II. Il soggetto nominativo come categoria filosoficamente carica

2.1 La struttura nominativo-accusativa e la sua ontologia implicita

Le lingue indoeuropee organizzano la struttura dei predicati verbali attraverso la distinzione tra soggetto nominativo e oggetto accusativo — una distinzione che la tradizione grammaticale ha formalizzato, a partire dai grammatici greci e latini, come la distinzione fondamentale della struttura della proposizione. Il soggetto è l'entità di cui si predica qualcosa; l'oggetto è l'entità che subisce l'azione del soggetto; il verbo esprime la relazione tra le due. Questa struttura, così naturalizzata da sembrare semplicemente la struttura della proposizione come tale, ha però delle proprietà specifiche che non sono universali tra le lingue del mondo e che hanno conseguenze filosofiche non banali.

La proprietà più rilevante per l'analisi che qui si conduce è quella che i tipologi linguistici chiamano «uniformità del soggetto»: in un sistema nominativo-accusativo, l'entità che ricopre la funzione di soggetto riceve lo stesso trattamento morfologico indipendentemente dalla natura della sua partecipazione all'evento. Il soggetto di un verbo intransitivo d'azione — Mario corre — il soggetto di un verbo transitivo — Mario costruisce la casa — il soggetto di un verbo di stato — Mario soffre — il soggetto di un predicato psicologico involontario — Mario si innamora — ricevono tutti la stessa marcatura nominativa. La morfologia non distingue: essa li livella tutti sotto la stessa categoria grammaticale. E poiché la struttura morfologica dell'enunciazione è il primo strumento attraverso cui la mente organizza l'esperienza in categorie

comunicabili, questo livellamento non è filosoficamente neutro: esso suggerisce, ad ogni atto di comunicazione, che vi sia qualcosa di comune e di fondamentale nella posizione di Mario in tutte queste situazioni diverse — che vi sia, cioè, un'unica funzione grammaticale «soggetto» che cattura adeguatamente la partecipazione dell'entità all'evento.

Questa suggestione non è innocua. Essa costituisce il substrato grammaticale su cui la tradizione filosofica ha costruito la nozione di «agente»: una categoria che si applica uniformemente a tutti i casi in cui un'entità è in posizione di soggetto nominativo, indipendentemente dalla qualità, dall'origine, dal grado di volontarietà o di controllo che caratterizza quella posizione. La filosofia ha poi cercato di differenziare internamente questa categoria — distinguendo tra azioni e accadimenti, tra azioni intenzionali e non intenzionali, tra agenti responsabili e non responsabili — ma lo ha fatto trattando queste distinzioni come distinzioni aggiuntive rispetto a una categoria base comune, non come distinzioni che ritagliano tipi fondamentalmente diversi di partecipazione a un evento.

2.2 La storia della nozione di agente nella tradizione analitica

È istruttivo tracciare brevemente come la nozione di agente sia stata elaborata nella tradizione analitica, perché il tracciato mostra con evidenza come la struttura grammaticale del soggetto nominativo abbia co-determinato la forma dei problemi.

Elizabeth Anscombe in *Intention* (1957) pone al centro della sua analisi dell'azione la domanda «perché?»: un evento è un'azione quando ha senso porgli questa domanda e ottenere risposta in termini di ragioni. La distinzione tra azione e accadimento è la distinzione tra ciò a cui si può rispondere con ragioni e ciò a cui non si può. Ma questa distinzione rimane binaria: o l'evento è un'azione (risposta in termini di ragioni possibile) oppure è un accadimento (risposta in termini di ragioni impossibile o inappropriata). L'analisi anscombeana non prevede una gradazione o una multidimensionalità dell'«essere-azione»: l'evento entra o non entra nella categoria.

Donald Davidson in «Actions, Reasons, and Causes» (1963) sviluppa la proposta causale: un evento è un'azione se è causato nella maniera appropriata dagli stati mentali dell'agente —

principalmente da una combinazione di desideri e credenze in una «pro-attitude». La proposta davidsoniana è influente e feconda, ma essa pure è binaria nella sua struttura: il causale nexus appropriato tra stati mentali ed evento o c'è o non c'è. Davidson non ha strumenti per distinguere tra un evento causato dagli stati mentali in modo completo (piena intenzione, pieno controllo) e un evento causato dagli stati mentali in modo parziale o mediato. Non che non si accorga di questi casi — la letteratura sull'akrasia e sulle azioni «non intenzionalmente intenzionali» ne è testimonianza — ma li tratta come variazioni entro una categoria, non come configurazioni distinte di dimensioni separate.

Harry Frankfurt in «Freedom of the Will and the Concept of a Person» (1971) introduce la struttura gerarchica dei desideri — il sistema di volizioni di primo e secondo ordine — come analisi della libertà della volontà. Questa proposta è più elaborata delle precedenti, ma rimane strutturalmente binaria: o vi è coerenza tra i livelli gerarchici della volontà (e allora vi è libertà) oppure vi è conflitto o incoerenza (e allora vi è coazione o debolezza della volontà). Le dimensioni dell'agentività che qui si intende distinguere — originazione causale, intenzionalità, controllabilità, responsabilità attribuibile — non compaiono come dimensioni distinte in Frankfurt: esse sono fuse nella struttura unica della volontà gerarchica.

Ciò che accomuna questi tre contributi — tra i più influenti della filosofia analitica dell'azione — è la struttura binaria dell'agentività: l'azione o è o non è nel senso pieno, l'agente o è o non è responsabile. Le distinzioni che vengono introdotte — tra tipi di causazione, tra livelli della volontà, tra diversi gradi di intenzionalità — operano come distinzioni secondarie rispetto a una categoria di base che rimane unitaria. E questa categoria di base è esattamente la categoria grammaticale del soggetto nominativo, che tratta come fondamentalmente identica la posizione di Mario in «Mario costruisce», «Mario starnutisce», «Mario soffre», «Mario si innamora».

2.3 L'analisi linguistica come strumento filosofico: un'avvertenza necessaria

Il principale rischio metodologico dell'approccio adottato in questo testo è quello dell'abuso dell'analogia linguistica: il rischio di costruire argomentazioni filosofiche che sembrano solide perché supportate da dati linguistici, ma che in realtà commettono errori di inferenza nel passaggio dai fatti

grammaticali alle conclusioni filosofiche. È necessario essere espliciti su questo rischio e sulle misure metodologiche adottate per contenere.

Il rischio specifico, nel caso presente, è il seguente: dal fatto che il georgiano grammaticalizza quattro dimensioni distinte dell'agentività non segue logicamente che queste quattro dimensioni siano le dimensioni naturali o necessarie dell'agentività umana. Potrebbero essere tre, o sei, o potrebbero non essere dimensioni distinte ma gradi di una scala unica. La scelta del georgiano come modello potrebbe essere arbitraria — potrebbe essere che il georgiano grammaticalizzi una partizione dell'agentività che riflette contingenze storiche o culturali della Georgia caucasica piuttosto che strutture universali dell'agentività umana. E il fatto che molte lingue condividano sistemi analoghi — il basco, le lingue caucasiche settentrionali, alcune lingue tibeto-birmane — non esclude questa possibilità, perché queste lingue potrebbero condividere la stessa contingenza storico-culturale.

Contro questo rischio si adottano due misure. La prima è la difesa indipendente della rilevanza filosofica di ciascuna delle quattro dimensioni: non si afferma semplicemente che il georgiano le distingue e quindi esse esistono, ma si mostra che ciascuna di esse è indipendentemente motivata da ragioni filosofiche che emergono dall'analisi dell'esperienza morale ordinaria e dai dibattiti della tradizione analitica stessa. Il georgiano funziona come risorsa euristica — come strumento che rende visibili distinzioni altrimenti nascoste — non come fondamento dell'argomento. La seconda misura è la consapevolezza esplicita che le quattro dimensioni qui proposte non sono necessariamente le dimensioni «giuste» o «complete»: esse sono le dimensioni che l'analisi del sistema georgiano suggerisce, e che si difendono sulla base di ragioni filosofiche indipendenti. Un'analisi di altra lingua o un'analisi filosofica di altra impostazione potrebbe suggerire un sistema diverso, e questo non invalidherebbe l'argomento principale — che è che la struttura binaria delle lingue indoeuropee lascia indistinte distinzioni filosoficamente rilevanti — ma modificherebbe il quadro delle distinzioni alternative proposte.

III. Il programma della monografia

3.1 Struttura e dipendenze tra le parti

La monografia procede nel modo seguente. La Parte Seconda presenta l'analisi del sistema morfologico georgiano con il grado di dettaglio linguistico necessario per fondare le affermazioni filosofiche successive. Non si tratta di una presentazione sommaria a fini divulgativi, ma di un'analisi abbastanza tecnica da resistere all'obiezione che il georgiano venga mal descritto o che la sua morfologia venga forzata per servire le tesi filosofiche. Questa sezione fa riferimento costante alla letteratura linguistica specialistica e si espone esplicitamente alle obiezioni dei linguisti caucasologi.

La Parte Terza sviluppa la critica alla tradizione analitica come lettura filosofica delle presupposizioni grammaticali. Essa prende le mosse dall'analisi dei testi di Anscombe, Davidson e Frankfurt, mostrando dove la struttura binaria dell'agentività opera come presupposizione implicita e come essa sia tracciabile, in modo storico-testuale, all'influenza del soggetto nominativo indoeuropeo. Questa sezione si muove nel dominio del secondo tipo di argomento delineato sopra — quello storico-ricostruttivo — e presenta le sue conclusioni come conclusioni interpretative, soggette a revisione sulla base di letture alternative dei testi.

La Parte Quarta è dedicata alla rilettura di Strawson e al problema degli atteggiamenti reattivi: essa costituisce il nucleo argomentativo più originale del testo, perché è qui che la critica grammaticale produce conseguenze più dirette e più difficili da assorbire nella letteratura esistente. L'argomento che gli atteggiamenti reattivi descritti da Strawson siano già pre-strutturati dalla grammatica binaria dell'agentività, e che una tassonomia più differenziata di quegli atteggiamenti sia sia possibile sia necessaria per un'analisi adeguata della responsabilità morale, viene sviluppato in questa sezione con il dettaglio filosofico che merita.

La Parte Quinta affronta la decomposizione del dibattito compatibilismo/incompatibilismo. Essa non pretende di risolvere il dibattito — pretesa che sarebbe infondata — ma di mostrare come la struttura del dibattito cambi quando si adottano le distinzioni dimensionali sviluppate nelle sezioni precedenti. Il risultato è una chiarificazione di dove il dibattito è produttivo e dove esso gira a vuoto per via della presupposizione binaria. La Parte Quinta è strutturalmente dipendente dalla Parte Quarta: le sue argomentazioni presuppongono le distinzioni sviluppate nell'analisi degli atteggiamenti reattivi.

3.2 Ciò che questa monografia non fa

È metodologicamente corretto essere espliciti sulle limitazioni dell'argomento, per prevenire obiezioni che non riguardano le tesi effettivamente difese.

Questa monografia non sostiene il relativismo filosofico. Non si afferma che le tradizioni filosofiche sviluppate in grammatiche diverse producano concezioni incommensurabili dell'agentività, né che la verità morale dipenda dalla struttura grammaticale della propria lingua. Si afferma che certe grammatiche rendono più difficile la formulazione di certe distinzioni, non che quelle distinzioni siano inaccessibili o che la loro assenza in una tradizione implichi la falsità di quella tradizione.

Questa monografia non pretende di derivare la filosofia dalla linguistica. Il georgiano non è la fonte delle conclusioni filosofiche qui difese: è una risorsa euristica che rende visibili distinzioni altrimenti nascoste, e che costituisce evidenza del fatto che quelle distinzioni sono grammaticalmente realizzabili in modo coerente. Le conclusioni filosofiche si difendono sulla base di ragioni filosofiche indipendenti dalla morfologia georgiana.

Questa monografia non argomenta che i parlanti del georgiano possiedano, in virtù della loro grammatica, intuizioni morali più sofisticate di quelle dei parlanti di lingue indoeuropee. La relazione tra struttura grammaticale e intuizioni morali è troppo complessa e troppo mediata da altri fattori culturali perché questa inferenza sia legittima. Ciò che si argomenta è più modesto e più preciso: che il sistema morfologico georgiano ha reso disponibile, come strumento euristico per la filosofia, un quadro concettuale dell'agentività più differenziato di quello disponibile nelle lingue indoeuropee.

Sezione Espansa — La Questione della Relatività Linguistica: Storia, Metodo, Implicazioni

Il dibattito Sapir-Whorf e la sua ricezione

Il quadro storico

Benjamin Lee Whorf morì nel 1941 a quarantaquattro anni, lasciando una serie di articoli e manoscritti che sarebbero stati pubblicati postumi a cura di John B. Carroll nel volume *Language, Thought, and Reality* del 1956. La ricezione di quel volume seguì un percorso che è diventato quasi paradigmatico nella storia delle idee scientifiche: una prima fase di entusiasmo acritico, in cui le tesi di Whorf vennero interpretate nelle loro versioni più forti e più drammatiche; una seconda fase di critica sistematica, in cui quelle stesse versioni forti vennero smontate empiricamente con una facilità che sembrava risolvere la questione una volta per tutte; e una terza fase — quella in cui ci troviamo ancora, sia pure con crescente chiarezza metodologica — in cui si cercano di recuperare versioni più moderate e più precise delle intuizioni di Whorf, separando ciò che è difendibile da ciò che era insostenibile.

Il contributo di Whorf che ha maggiormente influenzato la discussione successiva è la sua analisi della grammatica hopi, in cui egli sosteneva che la lingua degli Hopi — un popolo indigeno dell'Arizona sudoccidentale — grammaticalizzasse una concezione del tempo radicalmente diversa da quella delle lingue europee standard. Whorf sosteneva che la grammatica hopi non distinguesse tra passato, presente e futuro nel modo in cui lo fanno le lingue indoeuropee, e che questa struttura grammaticale producesse una concezione del tempo come processo ciclico di manifestazione piuttosto che come dimensione lineare con un presente privilegiato come punto di riferimento. La conclusione — formulata da Whorf con una libertà speculativa che i suoi critici avrebbero poi attaccato duramente — era che i parlanti dell'hopi esperissero il tempo in modo strutturalmente diverso dai parlanti di lingue europee, e che la loro fisica e la loro metafisica implicite fossero corrispondentemente diverse.

La critica principale all'analisi hopi di Whorf è venuta dall'antropologa Ekkehart Malotki, che in *Hopi Time* del 1983 ha sostenuto, sulla base di uno studio linguistico molto più esteso di quello condotto da Whorf, che la grammatica hopi grammaticalizza riferimenti temporali molto più sistematicamente di quanto Whorf avesse riconosciuto, e che la

tesi della totale assenza di tempo nella grammatica hopi era empiricamente insostenibile. Questa critica è stata generalmente accettata come convincente, e ha contribuito a discreditarla non solo le tesi di Whorf sull'hopi ma la tradizione della relatività linguistica in generale.

Ma la critica di Malotki, per quanto filologicamente accurata rispetto alle affermazioni specifiche di Whorf sulla grammatica hopi, lascia intatta la questione metodologica più generale: se la grammatica di una lingua rende obbligatoria la specificazione di certe distinzioni concettuali in ogni atto comunicativo, questo ha o non ha conseguenze sulle operazioni cognitive automatiche dei suoi parlanti? Questa domanda è distinta dalla domanda sull'hopi, e la risposta alla seconda non implica logicamente alcuna risposta alla prima.

La ripresa sistematica: Levinson, Boroditsky, Slobin

La ricerca sulla relatività linguistica ha ripreso vigore negli anni Novanta con una metodologia molto più rigorosa di quella di Whorf. Stephen Levinson e il suo gruppo al Max Planck Institute for Psycholinguistics hanno condotto una serie di studi cross-linguistici sulla cognizione spaziale, mostrando che parlanti di lingue che grammaticalizzano la direzione assoluta — nord, sud, est, ovest, indipendentemente dall'orientamento del corpo — piuttosto che quella relativa — destra, sinistra, davanti, dietro rispetto al corpo — mostrano differenze sistematiche nei compiti di orientamento spaziale non linguistico. I parlanti di lingue con sistema assoluto usano il sistema assoluto anche in compiti senza componente linguistica esplicita, come ricordare la posizione di oggetti su un tavolo dopo essere stati ruotati di novanta gradi.

Lera Boroditsky ha condotto studi su come la grammatica del tempo verbale influenzi la concezione dello spazio-tempo, mostrando che parlanti del mandarino — che usa metafore verticali per il tempo («il mese di sopra» per «il mese scorso») oltre a quelle orizzontali standard — sono più veloci nell'elaborare relazioni temporali quando vengono prime con disposizioni verticali piuttosto che orizzontali. Questi effetti, sia pure di piccola entità, sono stati trovati in compiti non linguistici, il che suggerisce che la grammatica influenzi la cognizione in modo che va oltre il semplice processamento linguistico.

Dan Slobin, in una serie di studi sulla descrizione di eventi in movimento, ha mostrato che parlanti di lingue che grammaticalizzano il modo del movimento nel verbo — come l'inglese: «he ran into the house», dove «ran» esprime il modo — descrivono gli eventi di movimento includendo informazioni sul modo con frequenza significativamente maggiore di quanto non facciano i parlanti di lingue che grammaticalizzano la direzione nel verbo — come lo spagnolo: «entró corriendo», dove «entró» esprime la direzione. Questi risultati di Slobin hanno implicazioni importanti perché mostrano effetti cognitivi non nelle variabili target di un compito sperimentale progettato ad hoc, ma nella distribuzione spontanea dell'attenzione durante la descrizione di eventi: la grammatica sembra influenzare ciò a cui i parlanti prestano attenzione automaticamente, non solo come codificano ciò che hanno già notato.

Perché questo dibattito è rilevante per la monografia

L'analisi fin qui condotta della relatività linguistica è rilevante per la monografia in modo preciso ma limitato. Come già dichiarato, l'argomento principale non dipende dalla tesi causale-cognitiva nella sua forma forte: non si afferma che la struttura grammaticale delle lingue indoeuropee abbia causato nelle menti dei filosofi analitici l'incapacità di distinguere le dimensioni dell'agentività. Ma essa è rilevante per tre ragioni più sottili.

La prima è che la ricerca di Slobin sulla distribuzione spontanea dell'attenzione è il tipo di evidenza empirica più vicina alle intuizioni che motivano l'argomento di questo testo. Se la grammatica influenza ciò a cui i parlanti prestano attenzione automaticamente durante la descrizione di eventi, allora è plausibile che essa influenzi ciò a cui i filosofi prestano attenzione automaticamente durante la costruzione di teorie dell'azione. Non si ha bisogno della tesi forte — «la grammatica impedisce certi pensieri» — ma di quella più debole e più difendibile: «la grammatica orienta l'attenzione automatica verso certi aspetti degli eventi a discapito di altri, e questo ha effetti cumulativi sulla struttura delle teorie costruite in quella lingua».

La seconda ragione è che il dibattito sulla relatività linguistica ha prodotto un inventario di metodologie per distinguere affermazioni difendibili da affermazioni infondate, e questo

inventario è utile per calibrare le affermazioni di questo testo. In particolare, la distinzione di Slobin tra «thinking for speaking» — il pensiero che viene attivato specificatamente per la produzione linguistica — e il pensiero in senso più ampio è rilevante: Slobin argomenta che la grammatica influenzi primariamente il primo, e che gli effetti sul secondo siano più limitati e più incerti. Questo è un motivo aggiuntivo per non fare affidamento sulla tesi causale-cognitiva forte, e per concentrare l'argomento sull'influenza della grammatica sulla struttura delle teorie — un tipo di «thinking for speaking» di ordine superiore — piuttosto che sulla cognizione in generale.

La terza ragione è che il dibattito sulla relatività linguistica ha mostrato che gli effetti grammaticali sulla cognizione tendono a essere specifici di dominio: la grammatica spaziale influenza la cognizione spaziale, la grammatica temporale influenza la cognizione temporale, e così via. Questo suggerisce che la grammatica dell'agentività — se essa ha effetti cognitivi — li avrà specificamente sulla cognizione dell'agentività, non sulla cognizione in generale. E questo rende l'ipotesi più precisa e più falsificabile: si potrebbe disegnare uno studio che verifichi se parlanti di lingue che grammaticalizzano le dimensioni dell'agentività in modo multidimensionale abbiano intuizioni morali differenti dai parlanti di lingue con sistema binario, in compiti che richiedono di valutare responsabilità in situazioni multidimensionali.

La questione della circolarità: osservare la propria grammatica dall'interno

Il problema in forma precisa

L'obiezione metodologica più profonda all'approccio di questo testo è quella che si può chiamare il problema della circolarità epistemica: se la struttura grammaticale delle lingue indoeuropee ha pre-strutturato il pensiero dei filosofi analitici, non ha pre-strutturato anche il pensiero dell'autore di questo testo, che pensa in italiano? E se è così, come può un'analisi condotta dall'interno della propria grammatica vedere le presupposizioni di quella grammatica?

Questa obiezione non è nuova: essa è una versione filosofica del problema più generale dell'autoriflessività critica, che si pone per qualsiasi impresa critica che miri a vedere i propri

presupposti. La psicanalisi ha il problema di usare la psiche per analizzare la psiche; la sociologia della conoscenza ha il problema di usare la conoscenza per analizzare la conoscenza; la critica dell'ideologia ha il problema di usare categorie ideologicamente situate per smascherare l'ideologia. Tutte queste imprese hanno le proprie strategie per gestire il problema, e nessuna di esse consiste nel pretendere di eliminare la circolarità: essa è ineliminabile. Si tratta piuttosto di gestirla in modo metodologicamente onesto.

Nel caso specifico di questo testo, la strategia è quella già indicata: la triangolazione attraverso sistemi grammaticali alternativi. Non si pretende di osservare la propria grammatica da una posizione neutrale al di sopra di tutte le grammatiche — posizione che non esiste — ma di usare il confronto con grammatiche strutturalmente diverse come strumento per rendere visibili, per contrasto, le presupposizioni della propria. Questa strategia non elimina la circolarità: la descrizione del sistema georgiano che un parlante madrelingua italiano produce è inevitabilmente interpretativa e filtrata attraverso le categorie grammaticali italiane. Ma produce una visibilità parziale che è meglio di nessuna visibilità, e che può essere corretta e raffinata attraverso il confronto con analisi prodotte da prospettive grammaticali diverse.

La triangolazione come metodo

La triangolazione, come metodo per aggirare la circolarità epistemica, ha un'analogia illuminante nella storia della misurazione astronomica. Prima dell'invenzione dei moderni strumenti di misura a distanza, l'unico modo per determinare la distanza di un corpo celeste lontano era la parallasse: la differenza nell'angolo di osservazione dello stesso corpo misurata da due posizioni diverse sulla superficie terrestre. Nessuna delle due misurazioni era «assoluta» o «neutrale»: entrambe erano prese da dentro il sistema che si cercava di misurare. Ma la loro differenza — la parallasse — permetteva di calcolare la distanza in modo preciso e affidabile.

La triangolazione grammaticale funziona in modo analogo. La descrizione del sistema georgiano prodotta da un parlante italiano non è neutrale: essa proietta categorie italiane sul sistema georgiano, e rischia di distorcerlo. La descrizione dello stesso sistema prodotta da un parlante georgiano non è neutrale neppure essa: proietta categorie georgiane sulla propria grammatica, e rischia di non vederla come struttura specifica

ma come dato ovvio. Ma la differenza tra le due descrizioni — la parallasse grammaticale — produce informazioni che nessuna delle due, presa isolatamente, avrebbe potuto produrre. È in questa differenza che si localizzano le presupposizioni di ciascun sistema, rese visibili per contrasto.

Questo metodo non è epistemicamente perfetto: la parallasse grammaticale è meno precisa della parallasse astronomica, perché le differenze tra sistemi grammaticali sono più difficile da quantificare delle differenze angolari. Ma esso è il metodo migliore disponibile per un'impresa che è strutturalmente autoreferenziale, e la sua imperfezione non è una ragione per abbandonarlo ma per applicarlo con consapevolezza dei propri limiti.

Le lingue filosofiche artificiali e i loro limiti

Vi è una tradizione, che va da Leibniz alla lingua logica di Frege alla logica dei predicati del primo ordine, che ha cercato di costruire un linguaggio filosofico «puro» — un linguaggio in cui le presupposizioni grammaticali delle lingue naturali siano eliminate attraverso la formalizzazione. Questa tradizione ha prodotto strumenti analitici di straordinaria potenza, e ha permesso di identificare e risolvere una serie di problemi che il linguaggio naturale rendeva oscuri — il problema del non-essere in Frege e Russell, il problema della definibilità in Tarski, il problema dell'implicatura in Grice.

Ma la tradizione delle lingue filosofiche artificiali non ha eliminato le presupposizioni grammaticali: le ha trasposte a un livello più astratto, dove sono meno visibili ma non meno presenti. La logica dei predicati del primo ordine è costruita intorno alla struttura predicato-argomento, che è una versione formalizzata della struttura soggetto-predicato delle lingue naturali indoeuropee. Essa grammaticalizza la binarietà soggetto/oggetto nella forma della distinzione predicato unario/predicato binario (predicato a uno o due argomenti), e lascia indistinte le dimensioni dell'agentività esattamente come fa la struttura nominativo-accusativa delle lingue naturali da cui è derivata.

In altri termini: la logica formale ha ereditato le presupposizioni grammaticali del greco antico e del latino, attraverso Frege, che pensava in tedesco e formalizzava strutture che aveva imparato da quelle lingue. Non si afferma che la logica formale sia inutile o distorta: si afferma che essa non è lo strumento adatto per vedere le presupposizioni grammaticali che ha ereditato, perché

esse sono incorporate nelle sue strutture fondamentali. Per vederle è necessario uscire dal sistema formale e guardarlo dall'esterno — e guardarlo dall'esterno significa confrontarlo con sistemi grammaticali che operano diversamente.

La voce media greca e ciò che si è perso

La voce media come categoria grammaticale dell'agentività intermedia

Le grammatiche delle lingue indoeuropee antiche — greco, sanscrito, ittita — avevano una categoria morfologica che le lingue moderne hanno quasi completamente perduto: la voce media. Nelle lingue moderne si ha la voce attiva («Mario costruisce») e la voce passiva («la casa viene costruita da Mario»), con la diatesi riflessiva («Mario si lava») come caso intermedio. Ma il greco antico aveva la voce media come categoria morfologicamente distinta dalle altre, con significato non riducibile né all'attivo né al passivo né al riflessivo.

La semantica della voce media greca è complessa e dibattuta, ma i suoi usi principali rivelano una struttura comune: la voce media grammaticalizza azioni in cui il soggetto è contemporaneamente il punto di partenza e il punto di arrivo dell'evento, azioni in cui l'attività del soggetto ricade su se stesso o su qualcosa che appartiene alla propria sfera, azioni in cui il soggetto ha un coinvolgimento speciale nell'evento — non come agente puro che causa qualcosa di esterno, ma come entità che è al tempo stesso agente e interessata all'esito dell'evento.

La lista degli usi della voce media greca include: azioni riflessive in senso stretto (lavarsi, vestirsi, pettinarsi); azioni in cui il soggetto agisce nel proprio interesse o a proprio beneficio (comprarsi qualcosa, difendere le proprie ragioni); azioni mentali e cognitive in cui il soggetto è l'origine e il luogo del processo (pensare, ricordare, deliberare, volere in certi contesti); cambiamenti di stato interni (riscaldarsi, addormentarsi, innamorarsi in certi contesti); azioni reciproche tra soggetto e qualcosa di esterno in cui il soggetto è pienamente implicato (combattere, lottare, competere). Questa lista mostra che la voce media non è una categoria unitaria dal punto di vista semantico — essa copre situazioni molto diverse — ma che tutte le situazioni che copre condividono la caratteristica di

collocare il soggetto in una posizione intermedia tra il polo dell'agentività pura (attivo) e il polo della patiency pura (passivo).

La perdita della voce media e le sue conseguenze filosofiche

La perdita della voce media nel passaggio dal greco antico alle lingue romanze e germaniche moderne non è una semplice perdita morfologica: essa comporta la perdita di una risorsa grammaticale per esprimere tipi intermedi di agentività in modo non marcato. In greco antico, l'azione mentale deliberativa poteva essere espressa con la voce media senza che fosse necessario specificare ulteriormente la sua natura: la voce media incorporava già l'informazione che il processo era interno al soggetto e che il soggetto ne era sia l'origine che il luogo. In italiano, questa informazione deve essere espressa attraverso perifrasi o rimane implicita nel contesto.

Aristotele usa la voce media in modo sistematico per caratterizzare certi tipi di azioni e certi tipi di stati: la deliberazione (*bouleusis*), la scelta (*prohairesis*), il desiderio razionale (*boulesis*) sono espressi con forme medie o in costruzioni che hanno una struttura semantica media. Non si afferma che Aristotele avesse accesso a distinzioni che i filosofi moderni non possono avere: si afferma che la morfologia del greco rendeva disponibili distinzioni che la morfologia delle lingue moderne non rende disponibili, e che alcune delle difficoltà nell'interpretare il pensiero aristotelico sull'azione volontaria in traduzione moderna dipendono precisamente da questa perdita.

La tradizione di traduzione latina di Aristotele ha affrontato il problema in modo parziale: il latino aveva ancora tracce di forme deponenti — verbi che hanno morfologia passiva ma significato attivo — che preservavano alcuni usi della media greca. Ma il sistema deponente latino è molto meno sistematico della voce media greca, e le lingue romanze e germaniche hanno eliminato anche quello. Il risultato è che la tradizione filosofica moderna si è trovata a dover riconcettualizzare, attraverso perifrasi e distinzioni analitiche esplicite, distinzioni che la grammatica greca aveva reso morfologicamente automatiche.

Questo è un caso di argomento storico-ricostruttivo nel senso delineato metodologicamente: non si afferma che la perdita

della voce media abbia causato incomprensioni filosofiche, ma si mostra come essa abbia reso più difficile l'espressione di certe intuizioni sull'agentività intermedia — intuizioni che la filosofia analitica ha poi dovuto recuperare attraverso strumenti concettuali espliciti (come la distinzione frankfortiana tra desideri di primo e secondo ordine, o la distinzione anscombeana tra azione e accadimento) dove il greco poteva affidarsi a risorse grammaticali automatiche.

Il sanscrito e la tradizione grammaticale indiana

Panini e la coscienza metalinguistica

La tradizione grammaticale indiana, a partire da Panini (circa IV secolo a.C.), ha sviluppato un apparato descrittivo della grammatica sanscrita di straordinaria precisione e sofisticazione — un apparato che non ha paralleli nella tradizione occidentale fino a Saussure e agli strutturalisti del Novecento. La grammatica di Panini, l'*Ashtadhyayi*, descrive la morfologia del sanscrito classico attraverso un sistema di regole formali che ha qualche analogia con i sistemi di riscrittura della linguistica generativa moderna.

Ciò che è rilevante per la nostra analisi è che la tradizione grammaticale di Panini aveva una coscienza esplicita della relazione tra struttura grammaticale e struttura dell'esperienza: le categorie grammaticali del sanscrito non erano trattate come mere convenzioni formali ma come riflessi della struttura dell'esperienza e del mondo. In particolare, la nozione di *karaka* — le «relazioni causali» che un'azione intrattiene con i propri partecipanti — era elaborata con una finezza che non trova paralleli nella tradizione grammaticale occidentale.

I *karaka* paniniani includono: *karta* (l'agente, colui che agisce autonomamente), *karma* (l'oggetto che viene modificato dall'azione), *karana* (lo strumento attraverso cui l'azione viene compiuta), *sampradana* (il recipiente o beneficiario), *apadana* (il punto di partenza del movimento), *adhibikarana* (il luogo o il contesto). Questa lista non è semplicemente un inventario dei casi grammaticali del sanscrito: è un'ontologia dell'azione che distingue tra diversi modi in cui un'entità può partecipare a un evento, e che include distinzioni che la struttura nominativo-accusativa delle lingue indoeuropee occidentali non grammaticalizza sistematicamente.

In particolare, la distinzione paniniana tra *karta* come «agente che agisce da sé, per propria autonomia» e altri tipi di partecipanti è una distinzione tra l'originarietà causale nel senso proprio e altri tipi di coinvolgimento nell'evento. Il *karta* paniniano non è semplicemente «il soggetto grammaticale»: è l'entità che è il punto di partenza autonomo dell'azione, che agisce per propria spinta interna piuttosto che come strumento di un agente più fondamentale o come locus di un processo che ha altrove la propria origine. Questa è esattamente la distinzione tra la prima dimensione dell'agentività (originarietà causale) e le altre, formulata esplicitamente come distinzione grammaticale in una tradizione non europea e indipendente dalla tradizione caucasica del georgiano.

La tradizione grammaticale paniniana mostra dunque che il problema delle presupposizioni grammaticali dell'agentività non è specifico della tradizione analitica occidentale: è un problema che si pone per qualsiasi tradizione che sviluppi strumenti formali per descrivere la struttura delle proposizioni, e che diverse tradizioni hanno affrontato in modi che riflettono le strutture grammaticali delle lingue in cui si sono sviluppate. La tradizione indiana — sviluppandosi in sanscrito, una lingua con sistema di *karaka* più elaborato di quello delle lingue europee moderne — ha prodotto strumenti di analisi dell'azione più fini di quelli disponibili nella tradizione aristotelica latina.

La Morfologia Georgiana: Analisi Tecnica e Implicazioni Filosofiche

I. Il sistema verbale georgiano: introduzione tecnica

4.1 Premessa alla descrizione

La descrizione che segue si basa principalmente sulla grammatica descrittiva di Alice Harris (*Georgian Syntax*, 1981), sul lavoro comparativo di George Hewitt (*Georgian: A Learner's Grammar*, 2005 e *The Indigenous Languages of the Caucasus*, vol. 1), sulla tipologia dell'ergatività di R.M.W. Dixon (*Ergativity*, 1994), e sugli studi di morfologia funzionale di Bernard Comrie. Dove

le interpretazioni degli studiosi divergono — e nella morfologia georgiana questo accade con notevole frequenza, dato che il sistema è di straordinaria complessità — si segnala la divergenza e si indica quale interpretazione si adotta e perché.

Una precisazione preliminare di natura metodologica riguarda la differenza tra descrizione linguistica e interpretazione filosofica. Questo testo non pretende di proporre una nuova analisi linguistica del georgiano: adopera le analisi già disponibili nella letteratura specialistica come strumento per identificare distinzioni concettuali che hanno rilevanza filosofica. L'interprete filosofica delle descrizioni linguistiche è sempre una scelta, e si cercherà di essere trasparenti su dove la descrizione finisce e l'interpretazione comincia.

4.2 Le serie verbali e la marcatura del caso: struttura fondamentale

Il verbo georgiano è morfologicamente molto complesso: esso incorpora obbligatoriamente, nella sua forma morfologica concreta, informazioni che in lingue come l'italiano sono distribuite tra il verbo, il soggetto, gli oggetti e avverbi separati. La morfologia verbale georgiana marca simultaneamente: la persona e il numero del soggetto; la persona e il numero dell'oggetto diretto (quando presente); la persona e il numero dell'oggetto indiretto (quando presente); il tempo, l'aspetto e il modo; e — ciò che più interessa per l'analisi che qui si sviluppa — la «serie» verbale, ovvero il raggruppamento temporale-aspettuale-modale che determina la marcatura del caso sui partecipanti dell'evento.

Il sistema delle serie verbali del georgiano è stato descritto in modo diverso da diversi studiosi. Harris identifica cinque serie principali; Tschenkéli ne identifica di più; la grammaticografia georgiana tradizionale usa una classificazione in parte diversa. Ciò che è concordemente riconosciuto, e che ha rilievo per l'analisi filosofica, è il seguente fatto: la marcatura del caso sul soggetto grammaticale di un verbo georgiano non è fissa, ma varia a seconda della serie verbale. Un verbo che ha il soggetto al nominativo nella Prima Serie (che corrisponde approssimativamente al presente-futuro indicativo) avrà il soggetto al dativo nella Terza Serie (il perfetto-pluperfetto evidenziale), e un verbo che ha il soggetto al nominativo nella Prima Serie può avere il soggetto all'ergativo nella Seconda Serie (il passato aoristo).

Questo fenomeno — chiamato nella letteratura «case shift» o «case alternation» — è la caratteristica del georgiano che più direttamente interessa l'analisi dell'agentività. Esso non è un'irregolarità o un residuo storico: è un sistema organizzato che riflette distinzioni semantiche precise. Comprendere quali distinzioni semantiche produce il case shift è il compito principale di questa sezione.

4.3 Le classi verbali e la loro struttura semantica

Prima di analizzare il case shift, è necessario introdurre la distinzione tra classi verbali, perché è questa distinzione — più del case shift isolatamente considerato — che articola la tassonomia dell'agentività rilevante per la filosofia.

I verbi georgiani si distribuiscono in classi morfologiche sulla base di proprietà semantico-tematiche. La classificazione più influente per l'analisi che qui si conduce è quella che distingue tra verbi di Classe I, II, III e IV, anche se questa classificazione è semplificatoria rispetto alla complessità effettiva del sistema e viene usata qui a scopi espositivi, con le precisazioni necessarie.

I verbi di Classe I sono verbi transitivi in cui il soggetto è un agente nel senso più robusto: un'entità che causa intenzionalmente e con controllo un cambiamento di stato nell'oggetto. I verbi che descrivono azioni manuali deliberate — costruire, scrivere, aprire, chiudere, lanciare — appartengono tipicamente a questa classe. Nella Seconda Serie (aoristo), il soggetto di questi verbi porta la marcatura ergativa; il soggetto dello stesso verbo nella Prima Serie porta la marcatura nominativa. Questa alternanza nominativo/ergativo nel corso del paradigma verbale è una delle caratteristiche più studiate del georgiano, e la sua interpretazione semantica è oggetto di dibattito.

I verbi di Classe II sono verbi intransitivi che descrivono mutamenti di stato dell'entità soggetto: cadere, crescere, rompersi, riscaldarsi. Il soggetto di questi verbi porta la marcatura nominativa nella Prima Serie e nella Seconda Serie. A differenza dei verbi di Classe I, i verbi di Classe II non hanno un agente nel senso tradizionale: l'entità soggetto è il paziente del mutamento, non il suo produttore.

I verbi di Classe III sono verbi transitivi e intransitivi in cui l'agentività è intermedia: essi descrivono spesso azioni fisiche non pienamente controllate o stati prodotti dall'agente ma non pienamente deliberati. La marcatura del caso del loro soggetto

varia in modi che hanno dato luogo a interpretazioni diverse in letteratura.

I verbi di Classe IV — la classe più rilevante per la filosofia della responsabilità e degli atteggiamenti reattivi — sono verbi che descrivono stati psicologici, emozioni, processi corporei involontari, stati cognitivi. Il soggetto di questi verbi porta la marcatura dativa nella Prima Serie: l'entità che «ha» l'emozione o lo stato è grammaticalmente il ricevente, l'obliquo, non il nominativo. La marcatura dativa sul soggetto dei verbi di Classe IV è uno dei fenomeni più discussi nella linguistica del georgiano, e la sua analisi semantica ha implicazioni dirette per la filosofia della mente e della responsabilità.

4.4 Il dativo come marcatura dell'esperienza non-agentivo

La costruzione dativa per i predicati psicologici in georgiano non è isolata: essa si trova in molte lingue del mondo, e la sua analisi ha generato una letteratura tipologica ampia. La distinzione fondamentale che la costruzione dativa grammaticalizza è quella tra l'**agente** (chi causa l'evento) e l'**esperimente** (chi sperimenta lo stato senza causarlo). In italiano questa distinzione si realizza a livello lessicale — attraverso la scelta del verbo — ma non a livello morfologico: il soggetto di «Mario costruisce» e il soggetto di «Mario soffre» ricevono la stessa marcatura nominativa. In georgiano, la distinzione è morfologicamente obbligatoria: il soggetto di un verbo di Classe IV non può ricevere la marcatura nominativa senza produrre un enunciato agrammaticale.

Ciò che rende la costruzione dativa georgiana particolarmente rilevante per la filosofia dell'agentività è la granularità delle distinzioni che essa grammaticalizza all'interno del dominio degli stati psicologici. Non tutti i predicati psicologici ricevono la costruzione dativa in georgiano: vi sono predicati psicologici che prendono il nominativo — quelli in cui l'esperimente ha un maggiore grado di controllo o di originarietà causale rispetto allo stato. Questa variazione interna alla classe dei predicati psicologici grammaticalizza distinzioni che sono filosoficamente significative.

Alice Harris, nella sua analisi del georgiano, documenta la distinzione tra verbi che descrivono stati emotivi «imposti» al soggetto dall'esterno o dalla propria fisiologia — prendono il dativo — e verbi che descrivono stati in cui il soggetto ha un

grado di controllo maggiore o un grado di attivazione causale maggiore — prendono il nominativo o l'ergativo. Questa distinzione non è facilmente riducibile a una distinzione semplice volontario/involontario, perché la gradazione è più fine di così: ci sono stati involontari che prendono costruzioni nominative, e ci sono processi parzialmente volontari che prendono costruzioni dative. L'analisi di Harris suggerisce che la variabile rilevante sia una combinazione di originarietà causale, controllabilità e implicazione della struttura intenzionale dell'agente — esattamente le dimensioni che questo testo identifica come filosoficamente distinte.

4.5 L'ergativo e la sua semantica di originarietà causale

L'analisi dell'ergativo georgiano è uno degli argomenti più controversi nella letteratura caucasologica. La posizione tradizionale — seguita da Tschenkéli e da una parte della grammaticografia georgiana — è che l'ergativo nel georgiano sia semplicemente il marcatore del soggetto agentivo transitivo nel passato, senza semantica aggiuntiva rispetto al nominativo. Questa posizione è stata criticata da Harris e da altri linguisti funzionalisti, che hanno mostrato come l'alternanza nominativo/ergativo nel paradigma verbale georgiano non sia puramente morfosintattica ma corrisponda a distinzioni semantiche rilevanti.

La semantica dell'ergativo georgiano — nella misura in cui è ricostruibile dall'analisi morfologica e dalla comparazione tipologica — sembra gravitare attorno alla nozione di **originarietà causale**: il soggetto ergativo è il punto da cui l'evento emana, la sorgente da cui l'azione si irradia verso l'esterno, producendo un cambiamento nell'oggetto. Questa semantica è diversa dalla semantica del soggetto agentivo semplicemente inteso come «chi compie l'azione»: essa specifica il modo in cui il soggetto è implicato nell'evento — come sorgente causale — piuttosto che semplicemente che esso è implicato.

Dixon, nella sua tipologia dell'ergatività, distingue tra ergatività morfologica (la distribuzione dei marcatori di caso) e ergatività semantica (la distribuzione semantica delle nozioni di agentività e patiency). Egli mostra che queste due dimensioni spesso si disaccoppiano nelle lingue del mondo, e che le lingue cosiddette «split ergative» — tra cui il georgiano — sono casi in cui il disaccoppiamento è grammaticalmente sistematizzato. In

georgiano, l'ergatività morfologica si attiva in certe serie verbali ma non in altre, e questa attivazione grammaticalizza distinzioni semantiche che il sistema nominativo-accusativo semplice non grammaticalizza.

L'interpretazione filosoficamente rilevante di questa struttura è la seguente: il georgiano grammaticalizza la distinzione tra due modi in cui un'entità può essere la sorgente di un evento. Nel modo nominativo, l'entità è il soggetto dell'evento nel senso che l'evento si svolge nel suo frame spazio-temporale, con la sua implicazione, ma non necessariamente con la sua originarietà causale piena. Nel modo ergativo, l'entità è la sorgente causale da cui l'evento origina, nel senso che senza quella entità come punto di partenza causale l'evento non si sarebbe prodotto. Questa distinzione non corrisponde esattamente alla distinzione filosofica tra agire e accadere, né alla distinzione tra azione intenzionale e non intenzionale: essa è più specifica e più fondamentale di entrambe — è la distinzione tra essere la fonte causale dell'evento e essere il locus in cui l'evento si svolge.

4.6 La terza dimensione: il controllo nell'esecuzione

La dimensione della controllabilità — il grado in cui il soggetto può modulare, interrompere, ridirigere l'evento nel corso della sua esecuzione — è grammaticalizzata in georgiano attraverso un meccanismo diverso rispetto all'alternanza nominativo/ergativo/dativo. Essa si manifesta principalmente nella distinzione tra classi verbali, attraverso le proprietà morfologiche che determinano a quale classe un verbo appartiene, e nella distinzione tra forme verbali che indicano eventi completati e forme che indicano eventi in corso — una distinzione che in georgiano ha conseguenze sulla marcatura del caso che non ha paralleli nelle lingue indoeuropee.

La differenza rilevante è questa: la marcatura del caso in georgiano non dipende solo dal tipo semantico del verbo (classe verbale), ma anche dallo stato dell'evento rispetto al suo completamento (aspetto). Un evento in corso — non ancora completato, ancora modificabile, ancora sotto il potenziale controllo del soggetto — può ricevere una marcatura del caso diversa rispetto allo stesso tipo di evento nel momento del suo completamento, quando il controllo del soggetto cessa di essere rilevante perché l'evento ha già prodotto il suo effetto. Questa sensibilità dell'assegnazione del caso all'aspetto verbale è una

caratteristica delle lingue a ergatività condizionata dall'aspetto, della quale il georgiano è uno degli esempi più elaborati.

La rilevanza filosofica di questa caratteristica è la seguente: in georgiano, la posizione grammaticale del soggetto rispetto al predicato non è fissa attraverso il tempo dell'evento, ma varia con il suo stato d'avanzamento. Un evento avviato ma non completato situa il suo soggetto in una posizione morfologicamente diversa rispetto allo stesso evento nel momento del suo completamento. Questa variazione grammaticalizza una concezione dell'agentività in cui il controllo sull'evento non è un tratto fisso del soggetto ma una proprietà dinamica che varia con il progresso dell'evento: il soggetto «controlla di più» un evento in corso di quanto non lo «controlli» un evento completato, perché nell'evento completato l'effetto è già realizzato e il controllo è ormai fuori luogo.

4.7 La quarta dimensione: la responsabilità attribuibile e le costruzioni causative

La quarta dimensione — la responsabilità attribuibile, il fatto che il soggetto sia il referente appropriato delle pratiche di lode, biasimo, sanzione — non è grammaticalizzata in georgiano attraverso un meccanismo morfologico unico e identificabile, ma emerge dall'interazione tra le dimensioni precedenti e dai costrutti causativi, che in georgiano hanno una struttura particolarmente elaborata.

Il georgiano ha un sistema ricco di costrutti causativi che permettono di specificare il grado e il tipo di causazione che lega un soggetto a un evento. I costrutti causativi diretti — in cui il soggetto causa direttamente l'evento — sono morfologicamente distinti dai costrutti causativi indiretti — in cui il soggetto causa l'evento attraverso un intermediario. Questa distinzione non è una distinzione di grado (causazione forte versus causazione debole), ma una distinzione di struttura: la causazione diretta implica che il soggetto sia il punto da cui l'azione si produce senza mediazione; la causazione indiretta implica che il soggetto produca l'azione attraverso l'attivazione di un'altra catena causale.

La rilevanza di questa distinzione per la responsabilità attribuibile è immediata: chi causa direttamente un evento è attribuibile la responsabilità in modo diverso rispetto a chi lo causa indirettamente. Questa intuizione è comune in filosofia

del diritto e in filosofia morale, ma in italiano — come in inglese e nelle altre lingue indoeuropee — essa deve essere articolata perifericamente, attraverso clausole esplicite che specificano il tipo di causazione. In georgiano, la distinzione tra causazione diretta e indiretta è incorporata nella morfologia verbale stessa: la forma del verbo indica già, senza che sia necessario specificarlo esplicitamente, se il soggetto è il causatore diretto o il causatore mediato dell'evento.

4.8 Un esempio sintetico: «innamorarsi» in georgiano e le sue implicazioni

Per rendere concreta l'analisi fin qui svolta, è utile soffermarsi su un esempio specifico che la letteratura sulla grammatica georgiana cita spesso come caso paradigmatico del dativo soggetto: il verbo che corrisponde a «innamorarsi» o «amare» in senso involontario.

In georgiano, la costruzione più comune per esprimere lo stato di innamoramento usa il caso dativo sul soggetto esperiente e una forma verbale di Classe IV. La struttura grammaticale dell'enunciato colloca morfologicamente l'entità innamorata nella posizione dell'obliquo, del ricevente dell'evento, non nella posizione del soggetto nominativo. Questa costruzione grammaticalizza una precisa ontologia dell'amore romantico involontario: l'entità che si innamora non è la sorgente causale dell'innamoramento — non «produce» l'amore nel senso in cui costruisce qualcosa — ma è il luogo in cui l'amore si produce, il recipiente in cui si riversa un processo causale che ha la sua origine altrove (nella bellezza o nelle qualità dell'amato, nel momento, nella situazione).

Questa costruzione dativa è diversa, in georgiano, dalla costruzione usata per esprimere l'amore deliberato — la cura attiva, la scelta consapevole di mantenere e coltivare il sentimento. Forme verbali diverse, con strutture morfologiche diverse, grammaticalizzano la distinzione tra l'innamoramento come stato che accade al soggetto e il coltivare-l'amore come attività che il soggetto compie. In italiano, questa distinzione — pur concettualmente disponibile e filosoficamente significativa — non è grammaticalmente obbligatoria: «amare» copre entrambe le situazioni senza differenziare.

Ciò che questa differenza grammaticale illumina, sul piano filosofico, è una distinzione che la letteratura sulla responsabilità morale per gli stati emotivi ha trattato con

strumenti inadeguati. Harry Frankfurt, nella sua analisi della responsabilità per le proprie emozioni, distingue tra emozioni che si «identificano» con la struttura della volontà dell'agente — emozioni su cui l'agente ha «volizioni di secondo ordine» — ed emozioni che si manifestano come forze estranee alla volontà dell'agente. Questa distinzione frankfortiana è concettualmente analoga alla distinzione grammaticalizzata in georgiano tra il dativo dell'esperiente involontario e il nominativo dell'agente attivo, ma la formulazione frankfortiana è più complessa e più tecnica di quanto la struttura grammaticale del georgiano non richieda, perché deve essere costruita deliberatamente in un sistema grammaticale che non la rende automaticamente disponibile.

4.9 Basco, lingue caucasiche settentrionali, tibetano: la convergenza tipologica

Prima di passare all'analisi filosofica, è necessario collocare il georgiano nel contesto tipologico più ampio, per mostrare che le distinzioni grammaticali che esso articola non sono una peculiarità isolata di una singola lingua, ma un tipo di soluzione grammaticale che diverse famiglie linguistiche hanno sviluppato indipendentemente.

Il basco — la cui origine rimane tuttora non classificata — ha un sistema di marcatura dell'agentività che è strutturalmente diverso dal georgiano nei dettagli morfologici, ma analogo nella struttura fondamentale: l'ergativo basco marca il soggetto delle costruzioni transitive in modo tale da implicare semanticamente un alto grado di agentività causale, mentre l'absolutivo marca sia il paziente delle costruzioni transitive che il soggetto delle costruzioni intransitive, segnalando che in entrambi i casi l'entità non è nella posizione privilegiata della sorgente causale. Il basco aggiunge a questo sistema una morfologia verbale di accordo polipersonale — il verbo basco concorda simultaneamente con il soggetto, l'oggetto diretto e l'oggetto indiretto — che permette di codificare morfologicamente strutture di causazione distribuite tra più partecipanti all'evento.

Le lingue caucasiche nordorientali — avar, lak, darginiano, lezghiano e le altre lingue del Daghestan — hanno sistemi ergativi di notevole complessità, con sistemi di classi nominali che interagiscono con la marcatura dell'agentività in modi che non trovano equivalenti nelle lingue indoeuropee. Particolarmente rilevante per l'analisi qui condotta è il fatto che in molte di queste lingue la distinzione tra agentività causale

diretta e agentività mediata è grammaticalmente marcata in modo sistematico, attraverso classi verbali e morfologia di accordo che distinguono tra tipi di partecipazione causale all'evento.

Il tibetano classico e colloquiale ha sviluppato un sistema morfologico verbale in cui la distinzione tra azioni volontarie e involontarie è grammaticalmente obbligatoria: forme verbali diverse si usano per descrivere eventi che il soggetto compie deliberatamente e eventi che avvengono al soggetto o attraverso il soggetto senza piena deliberazione. Questa distinzione tibetana è articolata in modo analogo alla distinzione georgiana tra Classi verbali, ma attraverso un sistema morfologico diverso. La sua rilevanza è che essa si è sviluppata in una lingua associata a una tradizione filosofica e religiosa — il buddhismo tibetano — in cui la distinzione tra agentività volontaria e involontaria ha un ruolo centrale nella concezione della responsabilità karmiche e della responsabilità morale. Il tibetano offre dunque, in modo più diretto che il georgiano, un caso in cui la struttura grammaticale dell'agentività si è intrecciata con la tradizione filosofica che si è sviluppata in quella lingua.

La convergenza tipologica di queste lingue — georgiano, basco, lingue daghestaniche, tibetano — non è sufficiente a dimostrare che le distinzioni che esse grammaticalizzano siano universali o necessarie. Ma essa è sufficiente a mostrare che si tratta di soluzioni grammaticali indipendentemente sviluppate in risposta a strutture dell'esperienza umana che queste comunità linguistiche hanno trovato abbastanza rilevanti da grammaticalizzare. E questo è sufficiente per l'argomento filosofico che si intende sviluppare: non che le distinzioni siano universali, ma che esse siano filosoficamente motivate e grammaticalmente realizzabili in modo coerente.

4.10 Le quattro dimensioni come costruzione filosofica, non lettura diretta

Si deve essere espliciti, a conclusione di questa sezione tecnica, su un punto metodologico cruciale. Le quattro dimensioni dell'agentività — originazione causale, intenzionalità, controllabilità, responsabilità attribuibile — non sono leggibili direttamente dalla morfologia georgiana come quattro categorie grammaticali ben distinte e isolate. La morfologia georgiana è un sistema complesso in cui queste dimensioni si intrecciano, si sovrappongono e si influenzano reciprocamente in modi che

non corrispondono a quattro assi ortogonali indipendenti. L'analisi filosofica che qui si propone è una ricostruzione interpretativa di ciò che il sistema georgiano grammaticalizza, non una trascrizione diretta.

Questo è il punto in cui la separazione tra descrizione linguistica e interpretazione filosofica deve essere più esplicita. I linguisti descrivono il sistema georgiano in termini di classi verbali, serie temporali, marcatura del caso, accordo polipersonale: questi sono i termini tecnici adeguati per una descrizione linguistica. La traduzione di queste categorie tecniche nelle quattro dimensioni filosofiche è un'operazione interpretativa che si difende sulla base di ragioni filosofiche — la rilevanza di queste dimensioni per l'analisi degli atteggiamenti reattivi e del dibattito sul libero arbitrio — non sulla base di una corrispondenza uno-a-uno tra morfemi georgiani e concetti filosofici.

È possibile che un linguista georgianista obietti che la quadripartizione qui proposta forza il sistema georgiano in uno schema filosofico che non gli è proprio, e che una descrizione linguistica più accurata suggerirebbe una partizione diversa. Questa obiezione è in linea di principio legittima, e l'argomento filosofico di questo testo non dipende interamente dalla correttezza della quadripartizione specifica: dipende dalla tesi più generale che il sistema georgiano grammaticalizzi distinzioni dell'agentività che il sistema nominativo-accusativo indoeuropeo non grammaticalizza, e che queste distinzioni siano filosoficamente rilevanti. Questa tesi più generale è compatibile con partizioni alternative purché esse articolino distinzioni più fini di quelle disponibili nel sistema binario.

Le Presupposizioni Grammaticali della Tradizione Analitica e la Critica a Strawson

I. Anscombe, Davidson, Frankfurt: una rilettura

5.1 Anscombe e la domanda «perché?»

Elizabeth Anscombe apre *Intention* con una mossa che ha la semplicità apparente dei gesti fondativi: propone che la categoria di azione intenzionale si identifichi con quegli eventi ai quali è applicabile la domanda «perché?» nel senso della richiesta di ragioni. «Perché hai alzato il braccio?» — «Per salutare» è una risposta in termini di ragioni; «Perché il tuo cuore batte?» non ammette la stessa risposta. La distinzione è netta, e ha il vantaggio di essere fenomenologicamente motivata: c'è effettivamente qualcosa di diverso nel tipo di risposta che si aspetta alle due domande.

Ma si consideri cosa questa distinzione lascia fuori dal campo dell'analisi. «Perché hai starnutito?» può ricevere sia una risposta causale («perché avevo il pepe nel naso») che una risposta in termini di ragioni in senso indebolito («per liberarmi dal prurito»). «Perché ti sei innamorato?» può ricevere una risposta mista: cause e ragioni si intrecciano in un modo che la domanda anscombeana non risolve. «Perché hai tremato?» può ricevere risposte che mescolano la causazione fisiologica con il contesto e l'intenzione. La distinzione binaria anscombeana tra ciò che risponde alla domanda «perché?» e ciò che non vi risponde taglia il continuum dell'agentività in un punto che non è ovvio, e taglia in modo da lasciare all'interno della categoria «azione intenzionale» una serie eterogenea di casi che differiscono enormemente tra loro rispetto all'originarietà causale, al controllo e alla piena deliberatezza.

La formulazione anscombeana riflette la struttura del verbo inglese e del verbo italiano: entrambe le lingue grammaticalizzano la distinzione azione/accadimento attraverso la scelta lessicale del verbo e attraverso la struttura sintattica, ma non attraverso la morfologia del soggetto. Non vi è un segnale morfologico obbligatorio che distingua «Mario starnutisce» da «Mario costruisce»: entrambi hanno Mario al nominativo, entrambi usano la terza persona singolare del verbo. La struttura grammaticale tratta i due eventi come strutturalmente paralleli, e questa struttura parallela suggerisce — non dimostra, ma suggerisce — che la distinzione rilevante sia una distinzione binaria tra i due tipi.

Se Anscombe avesse costruito la propria analisi in georgiano, la morfologia del sistema verbale le avrebbe reso disponibile una struttura in cui la distinzione tra «Mario starnutisce» e «Mario costruisce» — e la distinzione tra entrambi e «Mario si innamora» — non deve essere costruita attraverso l'analisi della

domanda «perché?», ma è già incorporata nella forma morfologica del predicato. Non si afferma che questa disponibilità grammaticale avrebbe necessariamente prodotto un'analisi diversa: si afferma che la sua assenza ha reso la distinzione binaria anscombeana più naturale di quanto sarebbe stata in un sistema grammaticale più differenziato.

5.2 Davidson e la giusta causazione

L'analisi causale di Davidson — un evento è un'azione se causato «nel modo giusto» dagli stati mentali appropriati — ha il merito di chiarire la connessione tra azioni e stati mentali, e il demerito di lasciare l'espressione «nel modo giusto» senza analisi precisa. Davidson è consapevole di questo problema — lo chiama «the wayward causal chain problem» — e ha cercato di affrontarlo in scritti successivi, senza giungere a una soluzione generalmente accettata. Ma il problema delle catene causali deviate non è l'unico limite dell'analisi davidsoniana.

Il limite più rilevante per l'analisi che qui si conduce è che la proposta davidsoniana tratta la causazione dagli stati mentali come una condizione binaria: o il nesso causale appropriato è presente — e allora l'evento è un'azione — oppure non lo è — e allora non lo è. Ma la morfologia georgiana suggerisce che il nesso causale tra stati mentali e evento possa variare lungo dimensioni distinte: il nesso può essere forte rispetto all'originarietà causale ma debole rispetto all'intenzionalità; può essere pieno rispetto all'intenzionalità ma parziale rispetto al controllo; può essere diretto rispetto all'esecuzione ma mediato rispetto alla struttura causale più ampia che produce i propri stati mentali.

Il problema delle catene causali deviate è, in questa prospettiva, una manifestazione specifica di un problema più generale: quello dell'assenza di una distinzione tra dimensioni del nesso causale. Una catena causale «deviata» — il caso di chi, cercando di uccidere uno sconosciuto con la propria automobile, viene paralizzato dal nervosismo e causa un incidente nel quale lo sconosciuto muore accidentalmente — è deviata non semplicemente perché il nesso causale non è «del tipo giusto», ma perché rispetto a dimensioni specifiche dell'agentività la connessione tra l'intenzione e l'evento non è quella attesa: l'originarietà causale dell'agente è presente (l'incidente origina da lui), l'intenzionalità rispetto all'esito finale è presente (voleva la morte dello sconosciuto), ma il controllo sull'esecuzione e la struttura causale che connette intenzione a esito sono

radicalmente diverse da quelle dell'azione deliberata e controllata. Questa configurazione multidimensionale non ha nome nell'analisi davidsoniana, perché l'analisi non dispone delle dimensioni per nominarla.

5.2bis La deviance causale e il problema del controllo: Davidson riesaminato

L'analisi davidsoniana lascia aperto un problema che la letteratura successiva ha chiamato «il problema della deviance causale»: vi sono casi in cui gli stati mentali di un agente causano un'azione «nel modo sbagliato», eppure la causazione è presente. L'esempio standard: un alpinista vuole mollare la corda per sbarazzarsi del compagno, e questo pensiero lo rende così nervoso che la corda gli scivola accidentalmente dalle mani. L'azione è stata causata dallo stato mentale, ma non era intenzionale nel senso rilevante. Davidson riconosce il problema e propone di risolverlo specificando che la causazione deve avvenire «nel modo giusto», senza però riuscire a spiegare in che cosa consista tale modo giusto.

Il quadro multidimensionale permette di riformulare questo problema con maggiore precisione. Il caso della deviance causale è un caso in cui la dimensione dell'originarietà causale è alta — gli stati mentali dell'agente sono causalmente rilevanti — ma la dimensione del controllo nell'esecuzione è bassa o assente: l'agente non ha guidato l'evento verso il suo compimento secondo le sue intenzioni. Questa distinzione è esattamente quella che il sistema georgiano grammaticalizza attraverso la separazione tra la marcatura ergativa dei verbi di azione controllata e la marcatura dativa dei verbi in cui il soggetto sperimenta senza controllare. In georgiano, la deviance causale non sarebbe un caso limite problematico: sarebbe un caso in cui la morfologia forza una scelta tra due costruzioni, rendendo visibile la differenza tra il contributo causale alto e il controllo esecutivo basso.

Questo non significa che il georgiano «risolva» il problema della deviance causale: il problema filosofico — come specificare il tipo di connessione causale rilevante per l'azione intenzionale — rimane aperto. Ma significa che il georgiano rende visibile la struttura del problema in modo da permetterci di affrontarlo direttamente: invece di cercare una formula per il «modo giusto» di causazione entro un sistema binario, ci si può chiedere quali dimensioni dell'agentività siano soddisfatte nel caso paradigmatico e quali manchino nel caso della deviance.

La risposta — originarietà causale sì, controllo esecutivo no — è già una risposta filosofica parziale, resa disponibile dalla sola articolazione multidimensionale.

5.3 Frankfurt e la gerarchia della volontà

La proposta frankfortiana della gerarchia dei desideri è, tra le analisi analitiche dell'agentività, quella che si avvicina di più alla multidimensionalità suggerita dalla morfologia georgiana — ma che al tempo stesso illustra più chiaramente come la struttura grammaticale indoeuropea abbia limitato la portata di un'intuizione corretta.

Frankfurt distingue tra desideri di primo ordine (volere fare X) e desideri di secondo ordine (volere avere il desiderio di fare X, o volere non avere il desiderio di fare X). Un agente è libero, per Frankfurt, quando i propri desideri di primo ordine sono in coerenza con i propri desideri di secondo ordine: quando fa ciò che vuole, e vuole volere farlo. La mancanza di libertà può manifestarsi in due modi: come coazione (si fa ciò che non si vuole fare) o come debolezza della volontà (si fa ciò che si vuole fare ma non si vuole voler fare).

Questa proposta cattura qualcosa di reale: la distinzione tra un'azione che esprime la struttura profonda della propria volontà e un'azione che la tradisce. Ma essa non distingue tra dimensioni diverse che possono variare indipendentemente. Un agente con piena coerenza tra desideri di primo e secondo ordine può avere quella struttura di desideri causalmente determinata da condizionamenti che egli non ha scelto: Frankfurt ammette questo ma sostiene che sia irrilevante. Un agente può avere coerenza gerarchica dei desideri ma mancanza di controllo sull'esecuzione: i propri desideri coerenti producono un'azione che non si riesce a portare a termine nel modo previsto, per mancanza di abilità tecnica o per circostanze imprevedibili. Un agente può avere coerenza gerarchica e pieno controllo ma essere il risultato di manipolazione esterna sofisticata che ha prodotto la struttura stessa della propria volontà.

In tutti questi casi, la struttura frankfortiana non offre strumenti per distinguere. E la ragione strutturale di questa insufficienza è la stessa identificata per Anscombe e Davidson: la gerarchia frankfortiana è un sistema monodimensionale — la coerenza o incoerenza tra livelli della volontà — sovrapposto a una categoria di base che rimane binaria (l'agente ha o non ha la struttura gerarchica appropriata). Le dimensioni

dell'originarietà causale, dell'intenzionalità, del controllo e della responsabilità attribuibile non sono articolate come dimensioni separate ma fuse nella struttura unitaria della volontà gerarchica.

5.3bis L'identità pratica e il limite della gerarchia: una critica estesa a Frankfurt

La struttura gerarchica frankfortiana è vulnerabile a una critica che il quadro multidimensionale articola in modo particolarmente chiaro. Frankfurt tratta la coerenza tra desideri di primo e secondo ordine come condizione sufficiente per la libertà della volontà; ma questa coerenza può essere il prodotto di processi causali che l'agente non ha controllato in nessun senso rilevante. Il caso del servo ben condizionato — che vuole genuinamente servire perché il suo condizionamento ha prodotto desideri di secondo ordine che approvano i suoi desideri di primo ordine di servire — è già nell'arsenale critico della letteratura. Frankfurt risponde che la manipolazione esterna delle volizioni è irrilevante purché il risultato sia una struttura coerente che l'agente possa identificare come propria. Questa risposta è filosoficamente insufficiente, ma la sua insufficienza è difficile da articolare senza disporre di una distinzione concettuale tra la dimensione dell'originarietà causale — come si è formata la struttura della volontà — e la dimensione dell'intenzionalità — come la struttura della volontà orienta l'azione in un momento dato.

Il sistema georgiano grammaticalizza esattamente questa distinzione. I verbi di classe I — che richiedono l'ergativo — grammaticalizzano azioni in cui l'agente è originatore causale dell'azione in modo pieno, non soltanto esecutore di una struttura di desideri già data. I verbi di classe II grammaticalizzano azioni in cui c'è originarietà causale dell'agente ma il controllo nell'esecuzione varia. I verbi di classe III grammaticalizzano stati in cui l'agente è esperiente senza originarietà causale. Il servo condizionato si trova in una situazione ibrida che il georgiano renderebbe morfologicamente visibile: i suoi desideri di secondo ordine hanno alta intenzionalità nel momento dell'azione — egli vuole genuinamente ciò che fa — ma bassa originarietà causale rispetto alla formazione di quella struttura di desideri. Frankfurt collassa queste due dimensioni in una; il georgiano le tiene separate; il quadro multidimensionale mostra che la distinzione è filosoficamente necessaria.

L'implicazione pratica è rilevante per la teoria della responsabilità. Se si accetta che originarietà causale e intenzionalità siano dimensioni distinte, allora si deve concludere che il condizionamento sistematico dei desideri di un individuo — attraverso manipolazione, propaganda, strutture di incentivi coercitive — riduce la sua responsabilità anche quando la sua struttura di desideri è perfettamente coerente. Non perché manchino i desideri di secondo ordine, ma perché la dimensione dell'originarietà causale è stata compromessa da forze esterne. Questa conclusione è intuitivamente corretta — riteniamo che le vittime di condizionamento sistematico abbiano responsabilità attenuata — ma la struttura frankfurtiana non riesce a fondarla.

5.4 La presupposizione grammaticale nella tradizione: una ricostruzione storico-testuale

Come si è sviluppata storicamente la presupposizione binaria dell'agentività nella tradizione analitica? Tracciarne la storia richiede di andare oltre i singoli testi e di considerare come il problema dell'azione sia stato ereditato dalla grammatica filosofica della tradizione greca e latina.

Il termine greco che è all'origine della nozione analitica di azione è *praxis*, e la distinzione aristotelica tra *praxis* (azione) e *poiesis* (produzione) è la fonte più diretta della distinzione tra azione e accadimento che troviamo in Anscombe. Ma ciò che è meno notato è che il greco antico, pur essendo una lingua nominativo-accusativa come l'inglese e l'italiano, ha un sistema verbale con marcature di voce — attivo, medio, passivo — che introduce distinzioni dell'agentività più fini di quelle disponibili in inglese. La voce media greca, in particolare, grammaticalizza una classe di azioni in cui il soggetto è al tempo stesso agente e paziente, in cui l'azione «ricade» sul soggetto stesso, in cui la distinzione tra agire e ricevere l'azione è sfumata. Questa categoria grammaticale ha influenzato la tradizione filosofica greca in modi che la filosofia analitica — sviluppata principalmente in lingue senza voce media grammaticale — ha perso.

La traduzione del corpus aristotelico in latino ha comportato una prima semplificazione: il latino ha una morfologia verbale meno ricca di quella greca, e alcune distinzioni che il greco grammaticalizzava attraverso la morfologia devono essere rese in latino attraverso perifrasi o vengono semplicemente perdute. La successiva traduzione del corpus aristotelico e scolastico

nelle lingue nazionali europee ha comportato ulteriori semplificazioni nella stessa direzione. La tradizione analitica anglofona ha ricevuto il problema dell'azione in un sistema grammaticale in cui le risorse morfologiche per differenziare tipi di agentività sono al minimo rispetto alla storia della tradizione occidentale.

II. Strawson: una critica sistematica

6.1 La struttura di *Freedom and Resentment* riesaminata

La rilettura di «*Freedom and Resentment*» proposta in questo testo non si limita a segnalare lacune nell'analisi di Strawson: essa cerca di mostrare come le lacune siano sistematiche, non accidentali, e come esse siano tracciabili alla struttura grammaticale del sistema in cui Strawson pensa.

L'argomento centrale di Strawson — che la giustificazione delle pratiche di responsabilità risieda nella struttura degli atteggiamenti reattivi piuttosto che in una teoria metafisica della libertà — è robusto e difendibile. Il problema non è questo argomento, ma il modo in cui Strawson descrive gli atteggiamenti reattivi che sono al cuore del sistema. La descrizione che Strawson offre è, come si argomenterà, una descrizione di quegli atteggiamenti nella loro forma indifferenziata — così come si presentano a un parlante di una lingua che grammaticalizza l'agentività in modo binario — piuttosto che una descrizione di essi nella loro struttura più ricca e differenziata.

6.2 La «qualità della volontà» come termine oscuro

Il concetto centrale attorno a cui Strawson costruisce la propria analisi degli atteggiamenti reattivi è quello di «qualità della volontà» (*quality of will*). Gli atteggiamenti reattivi sono, per Strawson, risposte alla qualità della volontà manifestata da un agente nei nostri confronti o nei confronti di terzi: se un agente mostra buona volontà — cura, rispetto, considerazione — gli atteggiamenti reattivi positivi (gratitudine, affetto, stima) sono appropriati; se mostra mala volontà — ostilità, indifferenza, disprezzo — gli atteggiamenti reattivi negativi (risentimento, biasimo, indignazione) sono appropriati.

Questo schema è intuitivamente plausibile, ma nasconde una difficoltà che emerge non appena ci si chiede che cosa esattamente si intenda per «qualità della volontà». La volontà è un aggregato di disposizioni, intenzioni, tendenze, abitudini, stati motivazionali: dire che la «qualità» di questa volontà è buona o cattiva presuppone che tale qualità sia valutabile lungo una dimensione unitaria. Ma questa presupposizione è esattamente quella che la morfologia georgiana mette in discussione: la «qualità» di un atto rispetto alla volontà di chi lo compie varia lungo dimensioni diverse che non si riducono a una scala unica.

Un chirurgo che opera con piena competenza tecnica e buona intenzione, producendo un esito fatale per ragioni parzialmente al di fuori del suo controllo, ha una «qualità della volontà» alta rispetto a quali dimensioni? L'intenzione era ottima; il controllo sull'esecuzione era alto; l'originarietà causale dell'esito fatale era distribuita tra lui e fattori indipendenti; la responsabilità attribuibile è parziale e graduata. Il concetto strawsoniano di «qualità della volontà» non ha strumenti per articolare questa configurazione: deve scegliere tra il dire che la qualità della volontà era alta (e quindi gli atteggiamenti reattivi positivi sono appropriati) o bassa (e quindi quelli negativi lo sono), senza la possibilità di dire che era alta rispetto ad alcune dimensioni e bassa rispetto ad altre.

Un genitore che, per via di una propria struttura psicologica non pienamente consapevole, trasmette ai figli modelli di comportamento dannosi: la «qualità della volontà» di questo genitore è buona (non intende causare danni, ama i propri figli) ma la sua originarietà causale rispetto ai danni è presente, e il suo controllo potenziale su quella trasmissione — anche se non esercitato — sarebbe stato in linea di principio disponibile. Quale atteggiamento reattivo è appropriato nei confronti di questo genitore? Il sistema strawsoniano non offre strumenti per rispondere con precisione.

6.3 Le «scuse» strawsoniane come sistema inadeguato

Strawson articola il sistema delle condizioni che attenuano o sospendono gli atteggiamenti reattivi attraverso due categorie principali: le «scuse» in senso stretto (condizioni che modificano la nostra interpretazione dell'atto specifico senza sospendere l'atteggiamento reattivo verso l'agente come tale) e

le condizioni di sospensione generale (malattia mentale, infanzia, grave perturbazione).

Il sistema delle scuse in senso stretto include, nella descrizione di Strawson, condizioni come: l'ignoranza dei fatti rilevanti (non sapeva che stava calpestando il piede dell'altro); la costrizione fisica (è stato spinto); l'incidente (è scivolato); l'allucinazione (credeva di vedere qualcosa di diverso da quello che c'era). Tutte queste condizioni hanno in comune, nell'analisi di Strawson, il fatto di mostrare che l'agente non aveva mala volontà nei nostri confronti: le scuse funzionano mostrando che la «qualità della volontà» era diversa da quella che sembrava.

Ma questa analisi è inadeguata per due ragioni distinte. La prima è che essa tratta tutte le scuse come varianti di un unico schema — la scusa mostra l'assenza di mala volontà — senza distinguere tra le dimensioni diverse su cui le scuse operano. «Non sapeva» (ignoranza dei fatti) opera sulla dimensione dell'intenzionalità: mostra che l'atto non era intenzionale nel senso rilevante. «È stato spinto» opera sulla dimensione dell'originarietà causale: mostra che il soggetto non era la sorgente causale dell'evento. «Non riusciva a controllarsi» opera sulla dimensione del controllo: mostra che il soggetto non aveva controllo sull'esecuzione dell'atto. «È stato ingannato» opera su una combinazione di intenzionalità e originarietà: mostra che l'intenzione era presente ma la sua realizzazione era mediata da fattori indipendenti. Tutte queste scuse hanno effetti diversi sugli atteggiamenti reattivi appropriati — e tuttavia Strawson le raggruppa tutte sotto il cappello unitario dell'«assenza di mala volontà», perdendo le distinzioni.

La seconda ragione per cui il sistema delle scuse è inadeguato è che esso tratta le condizioni di attenuazione degli atteggiamenti reattivi come condizioni binarie: o la scusa «funziona» — e allora il risentimento cede — oppure non funziona. Nella realtà dell'esperienza morale, le attenuazioni degli atteggiamenti reattivi sono graduate e multidimensionali: si prova meno risentimento verso chi ha agito con intenzionalità ridotta rispetto a chi ha agito con piena intenzionalità, ma il risentimento non scompare; si prova risentimento di tipo diverso verso chi aveva scarso controllo rispetto a verso chi aveva pieno controllo ma ha agito con indifferenza. Questi gradi e questi tipi diversi di risposta reattiva non sono catturati dal sistema strawsoniano, perché quest'ultimo non dispone delle dimensioni che li articolano.

6.3bis La morfologia georgiana degli atteggiamenti reattivi: un esperimento di pensiero

Vale la pena condurre esplicitamente l'esperimento di pensiero che la prospettiva tipologica suggerisce: come cambierebbe la formulazione di Freedom and Resentment se Strawson avesse filosofato in georgiano? La domanda non è meramente retorica: essa identifica con precisione dove si trovano i punti di pressione concettuale che la grammatica inglese nasconde.

In georgiano, il verbo che corrisponde al risentire — nel senso del risentire-verso-qualcuno-per-qualcosa — richiederebbe una costruzione che specifichi, attraverso la marcatura dei casi, la struttura della relazione causale tra il soggetto che prova risentimento, l'oggetto del risentimento, e l'azione che lo ha provocato. Non sarebbe possibile dire semplicemente «X prova risentimento verso Y per Z» usando un unico schema grammaticale per tutti i casi: la morfologia verbale forcerebbe a specificare se Y ha agito con piena originarietà causale o con originarietà ridotta, se Y aveva controllo sull'azione o la subiva parzialmente, se la struttura causale era diretta o mediata. Queste specificazioni non sarebbero opzionali o stilistiche: sarebbero morfologicamente obbligatorie, esattamente come in italiano è obbligatorio specificare il genere di un sostantivo.

La conseguenza filosofica di questo esperimento di pensiero è la seguente: un Strawson che filosofasse in georgiano non potrebbe formulare il suo problema come una dicotomia tra l'atteggiamento reattivo pieno e l'atteggiamento oggettivo, perché la morfologia della sua lingua non permetterebbe questa semplificazione. Egli si troverebbe costretto — dalla propria grammatica, non dalla propria acutezza filosofica — a distinguere tra diversi tipi di risposta reattiva corrispondenti a diverse configurazioni dell'agentività. Il problema che nella tradizione analitica anglofona ha richiesto decenni di elaborazione post-strawsoniana per essere anche solo formulato in modo adeguato — il problema della struttura multidimensionale degli atteggiamenti reattivi — si porrebbe immediatamente, nella sua forma più articolata, a chiunque filosofasse in georgiano. La grammatica compensa ciò che la filosofia deve guadagnare faticosamente.

Questo esperimento di pensiero non dimostra che il georgiano sia superiore all'inglese per fare filosofia: dimostra che grammatiche diverse rendono disponibili, come strutture di

default, distinzioni diverse, e che la difficoltà o la facilità di certi problemi filosofici è distribuita asimmetricamente tra le tradizioni linguistiche. La domanda normativa — quali distinzioni si dovrebbero rendere disponibili come strumenti filosofici — è una domanda filosofica, non grammaticale. Ma la risposta a questa domanda normativa viene facilitata dalla consapevolezza di quali distinzioni certe grammatiche già grammaticalizzano e perché.

6.4 R. Jay Wallace e la «capacità razionale»: un'analisi dei limiti

R. Jay Wallace in *Responsibility and the Moral Sentiments* (1994) ha sviluppato l'impostazione strawsoniana in una direzione che introduce maggiore precisione senza però superare il limite strutturale della binarietà. La proposta di Wallace è che la condizione di appropriatezza degli atteggiamenti reattivi sia la **capacità razionale** dell'agente: un agente è appropriatamente soggetto agli atteggiamenti reattivi se e solo se possiede le capacità di comprendere e di rispondere alle esigenze morali — la capacità di riconoscere ciò che le ragioni morali richiedono e di governare il proprio comportamento in conformità ad esse.

Questa proposta ha il merito di identificare una condizione specifica e analizzabile per l'appropriatezza degli atteggiamenti reattivi, anziché lasciare il tutto alla nozione vaga di «qualità della volontà». Ma essa rimane strutturalmente binaria: o si possiede la capacità razionale rilevante (e si è soggetti agli atteggiamenti reattivi) oppure non la si possiede (e si è trattati con l'atteggiamento oggettivo). La capacità razionale di Wallace è una proprietà che si ha o non si ha — sia pure con gradi intermedi ammessi come eccezioni alle due posizioni estreme — non un fascio di capacità distinte che possono variare indipendentemente.

Ma la nozione di «capacità di rispondere alle ragioni» è essa stessa composita in un modo che Wallace non articola. Rispondere alle ragioni richiede: la capacità di percepire le ragioni (riconoscere che una certa situazione richiede una certa risposta); la capacità di formarsi un'intenzione in risposta ad esse; la capacità di eseguire l'intenzione attraverso un'azione controllata; e la capacità di mantenere questa connessione tra ragione-intenzione-azione in modo stabile nel tempo. Queste capacità possono variare indipendentemente: una persona può avere piena capacità percettiva delle ragioni ma difficoltà nell'esecuzione controllata (stati dissociativi, disturbi del

controllo degli impulsi); può avere piena capacità esecutiva ma difficoltà nella percezione delle ragioni morali (certi tipi di psicopatia funzionale); può avere piena capacità in situazioni ordinarie ma significativa riduzione in situazioni di stress estremo o di manipolazione sofisticata. Il sistema di Wallace non distingue tra questi casi perché tratta la «capacità di rispondere alle ragioni» come un'unità indivisibile, esattamente come il soggetto nominativo indoeuropeo tratta l'agentività come unità indivisibile.

6.5 T.M. Scanlon e il problema dell'imputazione

T.M. Scanlon, in *What We Owe to Each Other* (1998) e in *Moral Dimensions* (2008), ha proposto di riformulare il problema della responsabilità in termini di **imputazione**: un atto è imputabile a un agente quando riflette le sue attitudini in un senso che lo rende rilevante per la valutazione di quelle attitudini. L'imputazione scanloniana è una relazione tra un atto e la struttura delle attitudini dell'agente: un atto è imputabile quando «esprime» le attitudini dell'agente in un senso determinato.

La proposta di Scanlon è sofisticata e cattura qualcosa di importante sulla struttura della responsabilità morale interpersonale. Ma essa presenta un limite specifico che l'analisi georgiana illumina: la nozione di «esprimere le attitudini» è ambigua rispetto alle dimensioni dell'agentività. Un atto può esprimere le attitudini dell'agente in modi diversi: come espressione diretta di un'intenzione (l'atto è precisamente ciò che l'agente si proponeva di fare); come espressione indiretta attraverso una struttura causale mediata (l'atto è il risultato di una catena causale che origina dalle attitudini dell'agente ma non è stata pianificata nei dettagli); come espressione di una disposizione senza intenzione specifica (l'atto manifesta una disposizione dell'agente anche se non era intenzionale in quel momento); come fallimento di espressione (l'agente aveva attitudini diverse da quelle che l'atto manifesta, ma per debolezza di volontà o per mancanza di controllo ha compiuto l'atto comunque).

Scanlon non distingue sistematicamente tra queste modalità di espressione, e la sua nozione di imputazione rimane unitaria rispetto a esse. La ragione — e qui si torna alla tesi grammaticale del testo — è che la struttura del verbo inglese e delle lingue indoeuropee non offre strumenti morfologici per differenziare queste modalità: «l'atto esprime le attitudini

dell'agente» è una proposizione con lo stesso struttura grammaticale indipendentemente da quale delle modalità sopra descritte si intenda.

6.6 Susan Wolf e la «sanità razionale»

Susan Wolf in *Freedom Within Reason* (1990) propone una concezione della libertà come «capacità di agire in conformità con la Ragione» — una formulazione che intende la Ragione con la maiuscola, nel senso di una norma oggettiva di valutazione che l'agente capace può riconoscere e seguire. Il limite della proposta di Wolf rispetto all'analisi che qui si conduce è analogo ai limiti già identificati in Wallace e Scanlon: la «sanità razionale» è una condizione binaria — l'agente è o non è razionalmente sano nel senso rilevante — e non distingue tra le dimensioni dell'agentività che possono variare indipendentemente.

Vi è però nella proposta di Wolf un elemento che merita attenzione specifica, perché esso si avvicina alla dimensione della responsabilità attribuibile che qui si identifica come la quarta dimensione. Wolf sostiene che un agente possa essere moralmente responsabile anche di azioni che non avrebbe potuto fare diversamente — nel senso in cui le fa per incapacità di percepire la norma morale corretta — se quella incapacità è essa stessa un riflesso della sua struttura morale. Un agente «moralmente cieco» — che non riesce a vedere la sofferenza degli altri come ragione morale rilevante — è responsabile delle proprie azioni non perché avrebbe potuto agire diversamente (forse non poteva), ma perché la cecità morale è essa stessa una caratteristica della sua struttura di agente che lo rende il referente appropriato del biasimo.

Questa mossa di Wolf si avvicina alla dimensione della responsabilità attribuibile, nel senso che riconosce che essa può sussistere anche in assenza di pieno controllo e di piena intenzione. Ma Wolf la argomenta in modo diverso da come si argomenterebbe nel quadro multidimensionale: invece di riconoscere che la responsabilità attribuibile è una dimensione distinta dall'intenzionalità e dalla controllabilità, Wolf sostiene che in certi casi la mancanza di controllo e di intenzionalità è essa stessa imputabile all'agente. Il risultato è lo stesso, ma la struttura dell'argomento è diversa — e la struttura dell'argomento determina dove si possono formulare obiezioni precise.

6.7 La fenomenologia degli atteggiamenti reattivi: la ricchezza che manca

Si giunge ora al punto più costruttivo della critica a Strawson e ai suoi successori: la proposta di una fenomenologia differenziata degli atteggiamenti reattivi che tenga conto della multidimensionalità dell'agentività.

Il punto di partenza è un'osservazione fenomenologica: gli atteggiamenti reattivi nella vita ordinaria sono più differenziati di quanto la tassonomia strawsoniana riconosca. Vi sono stati emotivi e valutativi che la vita morale produce in risposta a configurazioni specifiche dell'agentività altrui, e che non sono riducibili né al risentimento pieno né all'atteggiamento oggettivo né a gradazioni intermedie di questi. Questi stati esistono — chiunque li riconosce nell'esperienza — ma non hanno nome nelle lingue indoeuropee, e la filosofia non li ha sistematicamente tematizzati.

Considerare alcune di queste configurazioni. C'è la risposta emotiva a un'offesa che si riconosce come originata dal soggetto, e che si riconosce come non pienamente intenzionale nel momento della sua esecuzione, ma che si sa provenire da una struttura psicologica del soggetto che avrebbe potuto essere diversa se il soggetto avesse fatto diversamente in qualche punto del passato. Non è risentimento pieno: l'azione non era pienamente intenzionale. Non è l'atteggiamento oggettivo: l'agente ha una storia e una struttura che lo rendono il referente appropriato di qualche tipo di risposta interpersonale. Non è semplice tristezza o delusione: c'è un elemento di risposta all'agentività dell'altro, non solo alla situazione. Questo stato — per cui non vi è una parola in italiano né in inglese — ha la propria fenomenologia, la propria struttura causale, le proprie implicazioni per il rapporto interpersonale. La mancanza di un termine non implica la mancanza del fenomeno: implica che il fenomeno non è stato sufficientemente analizzato.

C'è la risposta emotiva a un beneficio che si sa essere stato prodotto dall'agente non come azione intenzionalmente rivolta a noi, ma come effetto collaterale di un'azione motivata da altro, che tuttavia ha richiesto al benefattore controllo, competenza e qualche sforzo. Non è la gratitudine piena che Strawson descrive — il beneficio non era intenzionalmente rivolto a noi — ma non è semplicemente il riconoscimento di una fortuna neutrale: vi è qualcosa nell'aver ricevuto questo

beneficio che si riferisce all'agentività del benefattore, alla sua competenza e al suo sforzo, anche se non al suo orientamento verso di noi. Ancora una volta: questo stato esiste fenomenologicamente, non ha un termine, e la filosofia non lo ha analizzato sistematicamente.

C'è la risposta all'agente che ha prodotto un danno significativo come risultato di una struttura di incentivi che egli ha attivamente mantenuto e valorizzato, pur senza compiere nessun atto singolo intenzionalmente lesivo. Il manager che ha costruito una cultura aziendale del risultato a tutti i costi, e che non ha mai ordinato a nessuno di fare nulla di illegale o di dannoso, ma il cui sistema di incentivi ha sistematicamente prodotto comportamenti dannosi. Non è l'oggetto appropriato del risentimento classico — non ha compiuto nessun atto ostile intenzionale nei nostri confronti. Non è l'oggetto dell'atteggiamento puramente oggettivo — è un agente razionale con piena capacità, autore della struttura in questione. È l'oggetto di qualcosa che potremmo chiamare, in mancanza di termine migliore, indignazione strutturale: la risposta a un'agentività che ha prodotto danni non attraverso atti singoli intenzionali ma attraverso la costruzione e il mantenimento di strutture causali dannose.

6.8 Contro l'obiezione di irrilevanza: Strawson non ha bisogno di queste distinzioni

L'obiezione più naturale alla critica sviluppata fin qui è la seguente: Strawson non aveva bisogno di fare le distinzioni che qui si propongono perché il suo argomento non lo richiedeva. Il punto di «Freedom and Resentment» non è offrire una tassonomia completa degli atteggiamenti reattivi: è mostrare che la giustificazione delle pratiche di responsabilità risiede nella struttura di questi atteggiamenti e non in una teoria metafisica della libertà. Per questo scopo, una tassonomia approssimativa è sufficiente.

Questa obiezione è parzialmente corretta, ma non difende il sistema strawsoniano dai problemi identificati. È vero che Strawson non aveva bisogno di una tassonomia degli atteggiamenti reattivi per il proprio argomento principale: ma i successori di Strawson — Wallace, Scanlon, Wolf — hanno costruito proprio su quella tassonomia una teoria della responsabilità, ed è la teoria della responsabilità che risulta inadeguata quando si tenta di applicarla alla complessità effettiva dei casi morali. L'insufficienza non si vede

nell'argomento principale di Strawson — che rimane valido — ma nelle sue applicazioni alle questioni della responsabilità morale che la tradizione successiva ha sviluppato sul fondamento strawsoniano.

Vi è poi un secondo problema con questa obiezione: essa assume che la tassonomia approssimativa degli atteggiamenti reattivi di Strawson sia sufficiente per i suoi scopi, ma non considera che la scelta di trattare gli atteggiamenti reattivi come indifferenziati rispetto alle dimensioni dell'agentività è essa stessa una scelta filosoficamente significativa — non una semplificazione metodologicamente innocua. Se gli atteggiamenti reattivi sono la struttura fondamentale della psicologia morale, come Strawson sostiene, allora una descrizione inadeguata di quegli atteggiamenti compromette la possibilità di costruire su di essi una teoria della responsabilità adeguata. La semplicità della tassonomia non è un pregio metodologico: è una limitazione strutturale.

Sezione Espansa — Il Dibattito Post-Strawsoniano: Una Ricognizione Critica

La critica di Gary Watson e l'integrazione multidimensionale: un resoconto sistematico

Gary Watson ha identificato con precisione l'ambiguità fondamentale nel programma strawsoniano distinguendo la responsabilità come «aretaic assessment» — la valutazione del carattere, delle virtù, dei vizi di un agente — dalla responsabilità come «accountability» — il soggetto appropriato di risentimento, indignazione, richiesta di giustificazione. Il caso di Robert Harris — l'assassino brutale la cui storia biografica di abuso infantile estremo è narrata in dettaglio da Watson — mostra la tensione nel sistema strawsoniano: Harris è chiaramente il referente appropriato della valutazione aretica (il suo carattere è brutale, le sue azioni sono moralmente orrende), ma la sua storia causale riduce almeno in parte la sua responsabilità come accountability (il suo carattere è stato prodotto da fattori fuori del suo controllo).

Il quadro multidimensionale risolve questa tensione articolando le dimensioni che Watson lascia implicite. La dimensione dell'intenzionalità è alta in Harris: le sue azioni sono pienamente intenzionali nel momento della loro esecuzione. La dimensione del controllo nell'esecuzione è alta: egli ha il pieno controllo delle proprie azioni. Ma la dimensione dell'originarietà causale rispetto alla formazione del proprio carattere è compromessa: la struttura psicologica che orienta le sue azioni è il prodotto di forze causali che egli non ha controllato e che hanno operato in un'età in cui non disponeva delle capacità cognitive e volitiva necessarie per resistere. La dimensione della responsabilità strutturale è quindi bassa rispetto alla formazione del carattere, anche se alta rispetto alle azioni singole. Questo profilo dimensionale spiega sia l'orrore morale che Harris ispira — risposta alle dimensioni alte — sia la compassione che la sua storia produce — risposta alla dimensione bassa dell'originarietà causale del carattere.

La risposta del quadro multidimensionale a Watson non è che la tensione che egli identifica si dissolve: è che la tensione ha una struttura precisa che permette di calibrare le risposte reattive in modo corrispondentemente differenziato. Non si tratta di scegliere tra l'atteggiamento reattivo pieno e l'atteggiamento oggettivo: si tratta di adottare atteggiamenti che rispondono alle dimensioni alte dove sono alte — biasimo per le azioni intenzionali, richiesta di giustificazione, indignazione — e atteggiamenti che rispondono alle dimensioni basse dove sono basse — comprensione biografica, interesse per le strutture causali che hanno prodotto il carattere, politiche che prevengano quelle strutture. Questi atteggiamenti non si elidono a vicenda: si stratificano.

Gary Watson e le due facce della responsabilità

La critica interna al sistema strawsoniano

Gary Watson è stato tra i primi lettori di Strawson a riconoscere che la struttura del saggio del 1962 aveva un'ambiguità fondamentale non risolta, e che da questa ambiguità dipendevano alcune delle tensioni più durature nella letteratura successiva. Nel suo articolo «Two Faces of Responsibility» del 1996 — che è, insieme al saggio su «Responsibility and the Limits of Evil» del 1987, il contributo più acuto alla letteratura post-strawsoniana — Watson identifica ciò che chiama le «due

facce» della responsabilità morale: la responsabilità come valutabilità (*accountability*) e la responsabilità come attribuibilità (*attributability*).

La responsabilità come valutabilità — la *accountability* — è la dimensione che Strawson aveva al centro della propria analisi: essa riguarda la condizione in cui un agente è appropriatamente soggetto agli atteggiamenti reattivi, in cui si può lodarlo o biasimarlo, ringraziarlo o risentirsi di lui. Questa dimensione è interpersonale: si è valutabili nei confronti di altri che possono reagire ai propri atti.

La responsabilità come attribuibilità — la *attributability* — è una dimensione distinta: essa riguarda la condizione in cui un atto può essere attribuito all'agente come espressione della propria struttura valutativa o caratteriale, indipendentemente dalla questione se sia appropriato reagirvi con atteggiamenti reattivi. Watson mostra che queste due dimensioni non coincidono: ci sono atti che sono attribuibili a un agente — che esprimono la sua struttura caratteriale — senza che sia appropriato reagirvi con atteggiamenti reattivi pieni; e ci sono condizioni in cui gli atteggiamenti reattivi sono appropriati rispetto a certi aspetti di un atto senza che l'atto sia pienamente attribuibile all'agente.

Il caso di Robert Harris — il condannato a morte descritto da Watson nel saggio del 1987, cresciuto in condizioni di abuso sistematico che sembrano spiegare causalmente la struttura della sua personalità violenta — è il caso paradigmatico in cui le due facce della responsabilità si disaccoppiano. Harris è valutabile in senso pieno: i suoi atti mostrano una mancanza di cura per le vite altrui che giustifica gli atteggiamenti reattivi negativi più forti. Ma egli è anche, in un certo senso, il prodotto di una storia causale che non ha scelto, e questa storia suggerisce che i suoi atti siano meno pienamente attribuibili a lui come espressione della propria struttura valutativa autonoma: egli non ha «scelto» di essere la persona che è, almeno non nel senso in cui lo richiederebbe la piena attribuibilità.

Watson usa questo caso per mostrare che il sistema strawsoniano non può rendere conto di questa tensione: esso offre solo il risentimento pieno o l'atteggiamento oggettivo, senza spazio per la risposta che Harris effettivamente produce in lettori attenti — una risposta che include sia l'orrore morale per i suoi atti sia qualcosa di molto vicino alla compassione per la persona che è diventato. Watson riconosce che questa tensione non ha una soluzione semplice: essa segnala una

complessità reale nella struttura della responsabilità morale che qualsiasi teoria adeguata deve affrontare.

La connessione con il quadro multidimensionale

La distinzione di Watson tra *accountability* e *attributability* è filosoficamente vicina alla distinzione che questo testo propone tra responsabilità attribuibile (la quarta dimensione) e le altre tre. La *accountability* di Watson corrisponde approssimativamente alla responsabilità attribuibile nel senso interpersonale — essere il referente appropriato degli atteggiamenti reattivi. La *attributability* corrisponde a una combinazione delle altre dimensioni: un atto è attribuibile all'agente nella misura in cui origina causalmente da lui, è compiuto con intenzionalità che esprime la sua struttura valutativa, e riflette il suo controllo sull'esecuzione.

Ma Watson non articola sistematicamente queste dimensioni interne all'*attributability*, e questo è il limite del suo sistema. La tensione che egli identifica nel caso Harris — tra la valutabilità piena e l'attribuibilità ridotta — emerge, nell'analisi multidimensionale, come la seguente configurazione: Harris ha alta originarietà relativa (i suoi atti originano dalla sua struttura psicologica), alta intenzionalità nel senso prossimo (agisce deliberatamente e con piena coscienza), ma l'originarietà causale della sua struttura psicologica è distribuita tra lui e la storia causale che lo ha prodotto, e il suo controllo sullo sviluppo di quella struttura era limitato in modo che sfida la piena attribuibilità. La tensione che Watson descrive è dunque la tensione tra le dimensioni interne dell'*attributability*: tra l'alta intenzionalità prossima e la bassa originarietà causale profonda.

Il quadro multidimensionale risolve la tensione di Watson non eliminandola ma articolandola: mostra che essa non è una tensione tra due proprietà della responsabilità che non riescono a stare insieme, ma una configurazione specifica delle quattro dimensioni in cui alcune sono alte e altre basse, e che richiede atteggiamenti reattivi corrispondentemente differenziati — non il risentimento pieno né l'atteggiamento oggettivo, ma qualcosa di più preciso che il vocabolario ordinario non ha ancora parola per nominare.

La connessione con il contrattualismo di Scanlon: originarietà e responsabilità interpersonale

T.M. Scanlon, in «What We Owe to Each Other» (1998) e in «Moral Dimensions» (2008), ha sviluppato una teoria della responsabilità che è in dialogo esplicito con Strawson ma che si muove in una direzione che il quadro multidimensionale può illuminare ulteriormente. Scanlon sostiene che la responsabilità per un'azione dipende da ciò che egli chiama «il significato dell'azione»: dall'insieme degli atteggiamenti, delle intenzioni e delle strutture di valore che l'azione manifesta e che determinano come essa dovrebbe essere valutata nelle relazioni interpersonali. La questione non è soltanto se l'agente era causalmente responsabile dell'azione, ma quali atteggiamenti verso gli altri l'azione rivela, e quindi come la propria risposta interpersonale a quell'azione dovrebbe essere calibrata.

Il sistema multidimensionale è naturalmente compatibile con il progetto scanloniano, e lo arricchisce in un punto specifico. Scanlon riconosce che le attenuazioni della responsabilità sono graduali e differenziate — diverse circostanze attenuanti producono diversi tipi di giustificazioni parziali — ma non ha una struttura sistematica per classificare queste attenuazioni. Il quadro multidimensionale offre questa struttura: le attenuazioni si classificano secondo quale dimensione dell'agentività è ridotta e in che misura. Un'azione compiuta con piena originarietà causale ma ridotto controllo esecutivo ha un «significato» scanloniano diverso da un'azione compiuta con pieno controllo esecutivo ma originarietà causale attribuibile in parte a fattori esterni: il tipo di risposta interpersonale appropriata è diverso non solo in grado ma in qualità, e questa differenza qualitativa corrisponde alla differenza dimensionale del quadro multidimensionale.

C'è però un punto in cui la prospettiva grammaticale suggerisce una critica specifica a Scanlon. Nel suo resoconto dell'imputabilità (attributability), Scanlon distingue tra l'essere l'autore di un'azione e l'essere responsabili per le sue conseguenze. Questa distinzione è importante e corretta; ma Scanlon la tratta come se l'«essere autore» fosse una condizione unitaria, che si ha o non si ha. Il sistema georgiano mostra invece che l'autorialità di un'azione è essa stessa multidimensionale: si può essere autore in senso pieno (con alta originarietà causale, alta intenzionalità, alto controllo), in senso parziale (con alcune dimensioni alte e altre basse), o in senso strutturale (attraverso la costruzione di sistemi causali che producono effetti non direttamente intenzionati). Questa gradazione dell'autorialità non smentisce Scanlon: la completa, e la rende più precisa.

Il problema della moral luck

Nagel e il paradosso dell'azione

Nel 1979, Thomas Nagel pubblicò un saggio destinato a diventare uno dei testi più citati nella filosofia morale analitica: «Moral Luck». L'argomento di Nagel parte da un'apparente contraddizione nella struttura della nostra pratica di attribuzione della responsabilità. Da un lato, sembra intuitivamente corretto che le persone siano responsabili solo per ciò che dipende da loro — solo per ciò che è sotto il loro controllo. Dall'altro, la pratica ordinaria di attribuzione della responsabilità dipende sistematicamente da fattori che non dipendono dall'agente: dal risultato effettivo delle proprie azioni (chi guida ubriaco e investe un pedone è trattato diversamente da chi guida ubriaco e arriva a casa senza incidenti), dalle circostanze in cui ci si trova (chi avrebbe commesso crimini di guerra se fosse vissuto nella Germania nazista è trattato diversamente da chi li ha effettivamente commessi), e persino dalla costituzione caratteriale con cui si è nati.

Nagel chiama questa dipendenza dai fattori fuori del controllo dell'agente «moral luck» — fortuna morale — e osserva che essa è pervasiva in modo che sfida qualsiasi teoria della responsabilità che faccia del controllo la condizione necessaria dell'attribuzione. Se si prende sul serio il principio che la responsabilità richiede il controllo, quasi nessun agente è pienamente responsabile di quasi nessun atto, perché quasi ogni atto dipende in modo significativo da fattori fuori del controllo dell'agente.

La risposta al problema dalla prospettiva multidimensionale

Il problema della fortuna morale è stato ampiamente discusso nella letteratura successiva, con soluzioni diverse che tendono o a sacrificare il principio del controllo (ammettendo che la fortuna sia moralmente rilevante nonostante il disagio intuitivo che questo produce) o a ridefinire il controllo in modo da escludere la fortuna come fattore (sostenendo che ciò che conta è il tipo di controllo che l'agente aveva, non il risultato effettivo che la fortuna ha determinato).

Il quadro multidimensionale offre una prospettiva sul problema della fortuna morale che differisce da entrambe le alternative. Esso suggerisce che il problema emerga precisamente dalla

confusione tra le dimensioni dell'agentività: la fortuna influisce in modo diverso sulle diverse dimensioni, e il disagio che il problema genera è il sintomo di questa differenza non articolata.

La fortuna risultante — la differenza tra chi guida ubriaco e non investe nessuno e chi guida ubriaco e investe un pedone — riguarda principalmente la prima dimensione: l'originarietà causale. Chi ha investito il pedone è l'originatore causale della morte nel senso pieno; chi non ha investito nessuno non lo è. Ma le altre dimensioni — intenzionalità (entrambi avevano le stesse intenzioni, o la stessa mancanza di intenzione verso la sicurezza altrui), controllo (entrambi avevano lo stesso grado di controllo ridotto per via dell'alcol) e responsabilità attribuibile rispetto alla condotta — sono identiche nei due casi.

La distinzione tra fortuna risultante e le altre dimensioni dell'agentività permette di formulare una risposta al problema di Nagel che non richiede né di accettare interamente la fortuna morale né di negarla. Si può sostenere che la fortuna risultante sia moralmente rilevante rispetto alla prima dimensione dell'agentività — l'originarietà causale — senza che questa rilevanza si propaghi alle altre dimensioni. Il conducente ubriaco che ha investito il pedone è l'originatore causale della morte, e questo produce obblighi specifici (risarcimento, riparazione, riconoscimento del danno) che l'altro conducente non ha. Ma entrambi sono ugualmente responsabili rispetto alle dimensioni dell'intenzionalità e del controllo — entrambi hanno messo a rischio la vita altrui con la stessa mancanza di cura — e la responsabilità morale rispetto a queste dimensioni non varia in funzione della fortuna risultante.

Questa risposta non è disponibile nel sistema binario: in quel sistema, si deve scegliere se la fortuna risultante sia moralmente rilevante in senso pieno (e allora i due conducenti sono trattati diversamente anche rispetto alla responsabilità morale complessiva) o non lo sia affatto (e allora la differenza nel risultato non cambia nulla nella valutazione morale). Il sistema multidimensionale permette di rispondere «la fortuna risultante è rilevante rispetto a alcune dimensioni ma non ad altre», il che sembra corrispondere meglio alle intuizioni ordinarie: si capisce che il conducente che ha causato la morte abbia obblighi di risarcimento che l'altro non ha, e al tempo stesso si capisce che la valutazione della condotta — della mancanza di cura per la sicurezza altrui — non debba variare in funzione del risultato.

La fortuna circostanziale e la fortuna costitutiva

Nagel distingue, oltre alla fortuna risultante, altri tipi di fortuna morale: la **fortuna circostanziale** (la fortuna di non trovarsi nelle circostanze che avrebbero portato a commettere atti immorali) e la **fortuna costitutiva** (la fortuna di avere una struttura caratteriale che non tende verso atti immorali). Questi due tipi di fortuna pongono problemi diversi che il quadro multidimensionale illumina in modo specifico.

La fortuna circostanziale — il caso di chi non si è mai trovato nelle circostanze della Germania nazista, e che pertanto non ha mai avuto l'occasione di cedere o resistere alle pressioni verso la partecipazione ai crimini — riguarda principalmente la dimensione della responsabilità attribuibile. Si è responsabili dei propri atti, non degli atti che avremmo compiuto in circostanze controfattuali. Ma la domanda «sarei stato responsabile se mi fossi trovato in quelle circostanze?» è una domanda sulla disposizione della propria struttura caratteriale, che è responsabile nel senso della dimensione dell'originarietà causale: la propria struttura caratteriale è «mia» nella misura in cui origina da scelte e da processi di formazione che mi appartengono.

La fortuna costitutiva — il caso di chi ha avuto la fortuna di nascere con una struttura caratteriale che tende verso la bontà piuttosto che verso il male — è il tipo di fortuna che più direttamente sfida il sistema della responsabilità, perché riguarda la dimensione più profonda dell'originarietà causale: se la mia struttura caratteriale non dipende da me nel senso in cui la mia originarietà causale richiederebbe, allora la responsabilità per gli atti che essa produce è minata alla radice.

Il quadro multidimensionale offre la risposta seguente: la fortuna costitutiva è rilevante per la prima dimensione (originarietà causale profonda) ma non per le altre tre. Una persona con struttura caratteriale moralmente deficiente che agisce in modo dannoso è pienamente responsabile rispetto all'intenzionalità (agisce deliberatamente), al controllo (può modulare la propria condotta in risposta a ragioni), e alla responsabilità attribuibile nel senso interpersonale (è il referente appropriato degli atteggiamenti reattivi). Il fatto che quella struttura caratteriale abbia radici causali che risalgono al di là dell'agente — nella genetica, nell'educazione, nella fortuna — è rilevante per la domanda sull'originarietà profonda ma non implica l'esenzione dalla responsabilità nelle altre tre dimensioni. Questa è la risposta compatibilista al problema

della fortuna costitutiva, ma formulata in un modo che non richiede di negare la rilevanza della fortuna: essa è rilevante per una dimensione specifica, non per tutte.

La fortuna neurale e le basi biologiche dell'agentività: implicazioni per la fortuna costitutiva

Il problema della fortuna costitutiva — la fortuna che concerne il tipo di persona che si è, il proprio carattere, le proprie disposizioni — ha acquisito una dimensione nuova e pressante con i progressi della neuroscienze cognitive negli ultimi tre decenni. Si sa ora con ragionevole certezza che le strutture neurali che mediano il controllo degli impulsi, la pianificazione a lungo termine, l'empatia, e la sensibilità alle ragioni morali variano significativamente tra individui per ragioni in larga parte fuori del controllo degli individui stessi: genetica, storia di sviluppo prenatale e perinatale, trauma precoce, esposizione a sostanze tossiche nell'infanzia, struttura dell'attaccamento nei primi anni di vita. Le persone che commettono certi tipi di reati — in particolare i reati violenti e impulsivi — mostrano sistematicamente differenze nella struttura e nel funzionamento della corteccia prefrontale e delle strutture limbiche che sono in parte il prodotto di fattori causali che non hanno scelto.

Queste scoperte neuroscientifiche pongono la fortuna costitutiva in una forma particolarmente acuta. Non si tratta soltanto di dire, con Nagel, che il carattere è in parte il prodotto di fattori fuori del nostro controllo: si tratta di dire che le stesse capacità di controllo — la dimensione del controllo che il sistema georgiano grammaticalizza — variano tra individui per ragioni biologiche e biografiche fuori del loro controllo. Questo non è il regresso galeniano della self-origination: non richiede che ci si origini causalmente da soli. Ma crea una forma di asymmetria nella distribuzione del controllo che il quadro multidimensionale deve affrontare esplicitamente.

La risposta che il quadro multidimensionale suggerisce è la seguente: la fortuna neurale non elimina la dimensione del controllo dalla valutazione morale, ma la calibra. Un agente con corteccia prefrontale strutturalmente compromessa per ragioni biologiche non controllate non ha zero controllo esecutivo: ha ridotto controllo in circostanze specifiche e per impulsi specifici. Questa riduzione è una riduzione dimensionale parziale, non una cancellazione della dimensione. Le risposte reattive appropriate a questo agente devono essere calibrate su

questa riduzione: non il pieno risentimento che si addice a chi aveva pieno controllo e ha agito con indifferenza, non il puro atteggiamento oggettivo che si addice a chi non ha agentività affatto, ma qualcosa di più specifico — un'indignazione che riconosce l'agentività parziale, un biasimo che è anche preoccupazione per la struttura causale che ha prodotto l'azione, un'aspettativa di cambiamento rivolta non all'individuo in isolamento ma all'insieme delle condizioni che lo hanno formato.

Fischer e Ravizza: il meccanismo *reasons-responsive*

La proposta in dettaglio

John Martin Fischer e Mark Ravizza hanno sviluppato, in *Responsibility and Control: A Theory of Moral Responsibility* del 1998, la proposta compatibilista più elaborata e più sistematica nel periodo post-strawsoniano. Essa merita un'analisi dettagliata perché è la proposta che più si avvicina al quadro multidimensionale proposto da questo testo, e il cui confronto con quel quadro è filosoficamente più produttivo.

Fischer e Ravizza propongono che la condizione di responsabilità morale per un'azione sia che quell'azione sia prodotta da un **meccanismo proprio dell'agente** che sia **appropriatamente responsivo alle ragioni**. La condizione della «proprietà» del meccanismo — che Fischer e Ravizza chiamano *mechanism ownership* — richiede che l'agente abbia «preso a proprio carico» il meccanismo che produce le proprie azioni: che lo abbia riconosciuto come proprio, che si veda come la fonte delle proprie azioni attraverso quel meccanismo. La condizione della «responsività alle ragioni» richiede che il meccanismo sia in grado di rispondere alle ragioni in modo appropriato: che produca azioni diverse in circostanze in cui le ragioni per agire diversamente sono sufficientemente forti.

L'analisi di Fischer e Ravizza della responsività alle ragioni è sofisticata: essa distingue tra *reasons-recognition* (la capacità di riconoscere che certe ragioni sono ragioni per agire in un certo modo) e *reasons-reactivity* (la capacità di agire in conformità con le ragioni riconosciute). Un agente può avere buona capacità di riconoscere le ragioni morali ma difficoltà nel reagire ad esse: è il caso della debolezza di volontà, in cui si riconosce che le ragioni per astenersi sono più forti delle ragioni per agire, ma si

agisce ugualmente. Fischer e Ravizza sostengono che la responsività richiesta per la responsabilità sia *reasons-recognition* ma non necessariamente *reasons-reactivity* regolare: è sufficiente che il meccanismo sia in grado di reagire alle ragioni in alcune circostanze, non che lo faccia in tutte.

I limiti della proposta rispetto al quadro multidimensionale

La proposta di Fischer e Ravizza è implicitamente multidimensionale — essa distingue tra *mechanism ownership*, *reasons-recognition* e *reasons-reactivity* come condizioni distinte — ma non esplicita questa multidimensionalità come struttura teorica fondamentale. Il risultato è che la proposta rimane vulnerabile a obiezioni che il quadro multidimensionale risolve.

Il primo problema riguarda la condizione di *mechanism ownership*. Fischer e Ravizza richiedono che l'agente «si veda come la fonte delle proprie azioni» attraverso il meccanismo: questa condizione ha una struttura epistemica e riflessiva — riguarda l'autoconcezione dell'agente — che non corrisponde alla struttura causale dell'originarietà. Un agente manipolato sofisticatamente potrebbe «vedere se stesso come la fonte delle proprie azioni» attraverso un meccanismo che in realtà è stato impiantato dall'esterno: in questo caso, la condizione di *mechanism ownership* di Fischer e Ravizza sarebbe soddisfatta ma l'agente non sarebbe responsabile nel senso intuitivamente richiesto. Questo è il problema che i filosofi chiamano «manipulation problem», e Fischer e Ravizza lo riconoscono senza riuscire a risolverlo completamente all'interno del proprio sistema.

Il quadro multidimensionale offre una risposta: la condizione di *mechanism ownership* di Fischer e Ravizza cerca di catturare la prima dimensione dell'agentività (originarietà causale), ma lo fa attraverso un criterio epistemico (l'autoconcezione) piuttosto che causale. Il criterio causale — l'azione origina causalmente dalla struttura psicologica dell'agente nel senso georgiano del dativo/ergativo — è più robusto: esso non è soddisfatto nel caso del manipolato neuroscientificamente, anche se la sua autoconcezione lo porta a vedersi come la fonte dei propri atti.

Il secondo problema riguarda il rapporto tra *reasons-recognition* e *reasons-reactivity*. Fischer e Ravizza sostengono che la responsività alle ragioni richiesta per la responsabilità sia principalmente la *reasons-recognition*: è sufficiente che il

meccanismo possa riconoscere le ragioni, anche se non reagisce regolarmente ad esse. Ma questa posizione produce un problema: un agente con alta capacità di riconoscere le ragioni morali ma con sistematica incapacità di agire in conformità ad esse — per via di compulsioni, dipendenze, disturbi del controllo degli impulsi — soddisfa la condizione di Fischer e Ravizza ma sembra meno responsabile degli agenti che non hanno questa sistematica disconnessione tra riconoscimento e reazione.

Il quadro multidimensionale articola questa difficoltà con precisione: la distinzione tra *reasons-recognition* e *reasons-reactivity* corrisponde alla distinzione tra la seconda dimensione (intenzionalità — la capacità di formare intenzioni in risposta alle ragioni) e la terza dimensione (controllo — la capacità di eseguire quelle intenzioni attraverso azioni controllate). Un agente che ha alta intenzionalità ma basso controllo — che «sa» cosa dovrebbe fare e «intende» farlo ma sistematicamente non riesce a tradurre l'intenzione in azione controllata — non è nella stessa posizione di un agente che ha alta intenzionalità e alto controllo. La proposta di Fischer e Ravizza non distingue queste situazioni in modo adeguato.

Fischer, Ravizza e il georgiano: responsività alle ragioni come controllo dimensionale

Il meccanismo *reasons-responsive* di Fischer e Ravizza può essere riformulato in modo più preciso usando il quadro multidimensionale, e questa riformulazione risolve alcune tensioni interne alla proposta originale. Fischer e Ravizza distinguono tra responsività forte alle ragioni — il meccanismo avrebbe prodotto un'azione diversa se le ragioni fossero state diverse, in modo regolare e comprensibile — e responsività debole — solo in alcuni mondi possibili il meccanismo è sensibile alle ragioni. Essi richiedono la responsività debole come condizione sufficiente per la responsabilità, ma questa scelta è difficile da giustificare senza un criterio indipendente per quale grado di responsività sia sufficiente.

Il sistema georgiano suggerisce che la responsività alle ragioni non è una dimensione unica ma si scompone nelle stesse quattro dimensioni che strutturano l'agentività in generale. Un meccanismo può essere altamente responsivo alle ragioni sul piano dell'originarietà causale — le ragioni determinano quale intenzione viene formata — ma poco responsivo sul piano del controllo esecutivo — l'esecuzione dell'intenzione dipende da

fattori non governati dalle ragioni. O viceversa: il meccanismo esecutivo può essere molto sensibile alle ragioni — modulare finemente l'esecuzione in risposta a variazioni delle condizioni — mentre la formazione dell'intenzione può essere in parte automatica o abitudinaria, meno direttamente governata da ragioni esplicite. Queste variazioni dimensionali nella responsività alle ragioni producono configurazioni diverse di responsabilità, che il sistema binario di Fischer e Ravizza non distingue.

La distinzione georgiana tra i verbi di classe I — azioni intenzionali con piena originarietà causale — e i verbi di classe III — stati e processi involontari in cui il dativo segnala l'esperienza senza originarietà causale — corrisponde a due estremi del continuum di responsività alle ragioni. Ma ciò che il georgiano grammaticalizza, e che Fischer e Ravizza non articolano, è lo spazio intermedio: i verbi di classe II, che grammaticalizzano azioni che hanno originarietà causale del soggetto ma che il soggetto non controlla in modo diretto nell'esecuzione. Questo spazio intermedio è il luogo in cui si trova la gran parte delle azioni ordinarie — le abitudini, le risposte automatiche calibrate, le azioni di routine — e la filosofia dell'azione ha speso energia sproporzionata sui casi estremi (azione intenzionale piena versus stato involontario puro) ignorando il centro dello spazio.

Il compatibilismo di Dennett e la critica dalla varietà del controllo

«Elbow Room» e la ridefinizione del controllo

Daniel Dennett in *Elbow Room* (1984) ha sviluppato la difesa compatibilista più vivace e filosoficamente provocatoria del controllo condizionale. La strategia di Dennett consiste nel mostrare che la nozione di «potere fare diversamente» che l'incompatibilista esige — la possibilità di fare diversamente tenendo fissi tutti gli antecedenti causali — è non solo incoerente con il determinismo ma anche irrilevante per le pratiche di responsabilità e per la nozione di libertà che dovrebbe importare a un essere umano razionale.

Dennett usa una serie di esperimenti mentali e di analogie con sistemi artificiali — automi, macchine di Turing, sistemi di intelligenza artificiale rudimentale — per mostrare che il tipo di controllo che conta per la responsabilità è il controllo che un

sistema ha sulla propria condotta in risposta a inputs informativi, non il controllo su un'impossibile auto-determinazione assoluta. Una macchina scacchistica che gioca bene agli scacchi esercita controllo sulle proprie mosse nel senso che le sue mosse dipendono da una valutazione delle posizioni: la macchina avrebbe mosso diversamente se la posizione sulla scacchiera fosse stata diversa. Questo è il tipo di controllo condizionale che Dennett propone come base per la responsabilità.

La critica che il quadro multidimensionale muove a Dennett non è che la proposta sia sbagliata — essa cattura qualcosa di importante sulla dimensione del controllo — ma che essa riduce tutte le dimensioni dell'agentività al controllo condizionale, perdendo le distinzioni tra le altre dimensioni. La nozione dennettiana di controllo incorpora implicitamente l'intenzionalità (il sistema risponde a ragioni, non a impulsi casuali), l'originarietà relativa (il sistema produce le proprie risposte dalla propria struttura computazionale) e la responsabilità attribuibile (il sistema è il referente appropriato delle pratiche di miglioramento e correzione). Ma tutte queste dimensioni sono incorporate in una nozione unitaria di controllo, senza essere articolate come dimensioni distinte.

Il risultato è che il sistema dennettiano, come quello di Fischer e Ravizza, non riesce a distinguere adeguatamente casi in cui alcune dimensioni sono presenti e altre assenti. Un sistema che ha alto controllo condizionale — risponde alle ragioni in modo appropriato — ma bassa originarietà causale relativa — il controllo è il risultato di una manipolazione sofisticata che ha impiantato dall'esterno la struttura responsiva — è trattato da Dennett come responsabile nella misura in cui esercita il controllo condizionale. Ma questo sembra intuitivamente insufficiente: ciò che manca in questo caso non è il controllo, ma l'originarietà causale che fa sì che il controllo sia «del» soggetto nel senso rilevante.

Il problema della responsabilità collettiva

Un'estensione necessaria

Il dibattito sul libero arbitrio e sulla responsabilità morale si è concentrato quasi esclusivamente sulla responsabilità individuale, e ha trattato la responsabilità collettiva come una

questione derivata — come responsabilità di individui aggregati piuttosto che come responsabilità di collettivi come tali. Questa scelta metodologica è comprensibile da una prospettiva che individua nell'agentività individuale la forma primaria di agentività, ma essa riflette ancora una volta la presupposizione grammaticale del soggetto nominativo individuale: in italiano come in inglese, il soggetto grammaticale standard è un'entità individuale, e i collettivi come soggetti richiedono costruzioni grammaticali marcate o perifrasi.

Il quadro multidimensionale dell'agentività si applica alla responsabilità collettiva in modo che produce risultati interessanti. Una corporazione che produce sistematicamente danni all'ambiente attraverso le proprie pratiche operative può avere alta originarietà causale relativa (i danni originano dalle sue pratiche), alta «intenzionalità» nel senso delle sue politiche deliberate (le pratiche riflettono scelte consapevoli), ma controllo e responsabilità attribuibile distribuite in modo complesso tra i diversi livelli della struttura organizzativa.

La tradizione filosofica e giuridica ha faticato a trattare questa complessità perché la struttura nominativo-accusativa delle lingue in cui si è sviluppata grammaticalizza il soggetto come entità individuale. La responsabilità collettiva deve essere costruita artificialmente attraverso la *fiction iuris* della persona giuridica — che tratta il collettivo come se fosse un individuo — o attraverso la distribuzione della responsabilità ai singoli individui che compongono il collettivo. Entrambe le strategie perdono qualcosa della struttura effettiva dell'agentività collettiva.

Le lingue con sistemi di classi nominali che includono collettivi come entità grammaticalmente primitive — alcune lingue bantu, alcune lingue amerindiane — trattano i collettivi come soggetti grammaticali nel senso pieno, senza la necessità di costruirli artificialmente attraverso finzioni giuridiche. In queste lingue, la predicazione di originarietà causale, intenzionalità, controllo e responsabilità attribuibile può applicarsi direttamente al collettivo senza ridurlo alla somma dei suoi individui. Questo non è solo un'osservazione tipologica: è una risorsa filosofica per pensare la responsabilità collettiva in modo più adeguato alla sua struttura effettiva.

La Decomposizione del Libero Arbitrio

I. Perché il dibattito non progredisce

7.1 La struttura del problema e il suo stallo

Il dibattito tra compatibilismo e incompatibilismo ha una struttura peculiare: esso è allo stesso tempo tecnicamente raffinatissimo e praticamente immobile. Dopo secoli di confronto e dopo la produzione di una letteratura di straordinaria ricchezza, le posizioni fondamentali rimangono esattamente dove erano. Questa immobilità non è il segno di un problema genuinamente insolubile: è il segno di un problema mal formulato.

La formulazione standard del problema è: «L'azione libera è compatibile con il determinismo causale?» Questa formulazione presuppone tre cose che non vengono mai messe in questione dall'interno del dibattito. Prima, che «libera» sia un predicato unitario applicabile all'azione come insieme. Seconda, che il «determinismo» sia una tesi univoca con implicazioni chiare per la libertà. Terza, che la domanda di compatibilità abbia una risposta unica — sì o no — valida per tutti i tipi di libertà e per tutti i tipi di determinismo.

L'argomento di questo testo riguarda principalmente la prima presupposizione: il fatto che «libera» sia trattata come predicato unitario. Le altre due presupposizioni — l'univocità del determinismo e l'unicità della risposta di compatibilità — sono già state messe in discussione nella letteratura, sia pure in modo non sempre collegato al problema grammaticale. John Martin Fischer, ad esempio, ha mostrato che si possono distinguere tipi diversi di determinismo con implicazioni diverse per la libertà. David Lewis ha mostrato che la domanda controfattuale «avrebbe potuto agire altrimenti» richiede una specificazione precisa delle condizioni in cui il controfattuale è valutato. Ma nessuno ha mostrato sistematicamente che la prima presupposizione — l'unità del predicato «libera» — è essa stessa problematica, né ha tracciato questa problematicità alla struttura grammaticale del predicato nelle lingue in cui il dibattito si è svolto.

7.2 La storia del dibattito come storia di una presupposizione

È istruttivo tracciare come la presupposizione di unitarietà del predicato «libera» abbia attraversato il dibattito, producendo le sue conseguenze a ogni passaggio.

Hume, nel *Treatise* e nell'*Enquiry*, propone la distinzione tra libertà come assenza di costrizione esterna e libertà come assenza di determinazione causale. Per Hume, la prima è sia possibile che desiderabile in un mondo deterministico; la seconda è impossibile e anche — se fosse possibile — non sarebbe la libertà che importa per la responsabilità. Ma questa distinzione humiana, per quanto acuta, rimane binaria: c'è la libertà che conta (assenza di costrizione) e quella che non conta (assenza di determinazione). Le dimensioni multiple dell'agentività non emergono.

Kant, in senso opposto, introduce la causalità per libertà come tipo di causazione radicalmente diversa dalla causalità naturale: la libertà trascendentale è incompatibile con il determinismo proprio perché è un tipo diverso di causazione, non riconducibile alla catena causale dei fenomeni. Ma anche Kant tratta la libertà come predicato unitario — l'alternativa è tra causazione fenomenica (deterministica) e causazione noumenica (libera) — senza articolare le dimensioni interne di ciascuna.

Il compatibilismo contemporaneo — da Dennett a Fischer a Wolf — ha sofisticato molto la proposta humiana, introducendo distinzioni tra tipi di controllo, tra tipi di responsività alle ragioni, tra tipi di originarietà causale. Ma ogni volta che un compatibilista introduce una distinzione, l'incompatibilista risponde che quella distinzione non cattura la «vera» libertà che importa, e il compatibilista deve cercare un'altra distinzione. Questo scambio è non-convergente perché la domanda «qual è la vera libertà?» non ha risposta all'interno della struttura in cui è posta: ha risposta solo quando si riconosce che «libertà» non è un predicato unitario ma un nome per un insieme di proprietà distinte che possono variare indipendentemente.

7.3 Le quattro domande: formulazione precisa

La decomposizione del problema del libero arbitrio in quattro domande distinte che questo testo propone non è una mossa

arbitraria: essa è motivata dall'analisi delle quattro dimensioni dell'agentività sviluppata nelle sezioni precedenti. Le quattro domande sono:

Prima domanda: L'originarietà causale di un'azione — il fatto che essa emani dal soggetto come sorgente causale — è compatibile con il determinismo?

Seconda domanda: L'intenzionalità di un'azione — il fatto che essa sia compiuta per ragioni, con deliberazione, in modo coerente con la struttura della volontà dell'agente — è compatibile con il determinismo?

Terza domanda: Il controllo di un agente sulla propria azione — la sua capacità di modulare, interrompere, ridirigere l'azione nel corso della sua esecuzione — è compatibile con il determinismo?

Quarta domanda: La responsabilità attribuibile — il fatto che l'agente sia il referente appropriato delle pratiche di lode, biasimo, sanzione, ricompensa — è compatibile con il determinismo?

Queste quattro domande non sono varianti di una stessa domanda: sono domande distinte con strutture logiche diverse e con risposte che si motivano con ragioni diverse. Ciò che rende la loro separazione importante non è solo che esse abbiano risposte diverse — questo potrebbe essere vero anche di varianti di una stessa domanda — ma che abbiano strutture di dipendenza diverse rispetto al determinismo: le ragioni per cui il determinismo minaccia o non minaccia l'originarietà causale sono diverse dalle ragioni per cui minaccia o non minaccia l'intenzionalità, e così via.

II. L'analisi delle quattro domande

8.1 Originarietà causale e determinismo: l'incompatibilista ha una presa

La prima domanda — se l'originarietà causale sia compatibile con il determinismo — è quella su cui l'incompatibilista ha le ragioni più solide, e dove il compatibilista deve fare la concessione più significativa. L'argomento incompatibilista sull'originarietà causale può essere formulato nel modo seguente: in un mondo deterministico, ogni azione di ogni agente è l'effetto di cause che precedono la nascita di

quell'agente, e queste cause precedono a loro volta la nascita dei genitori di quell'agente, e così risalendo indefinitamente. Nessun punto della catena causale è «originato» dall'agente nel senso in cui emanare-da-sé-come-sorgente-causale sembra richiedere: ogni punto è già determinato da punti precedenti.

Questo argomento è solido, e il tentativo compatibilista di confutarlo sostenendo che l'originarietà causale relativa — l'azione origina dagli stati mentali dell'agente nel momento in cui viene compiuta — sia sufficiente non è completamente convincente. L'incompatibilista ha ragione che c'è qualcosa di insoddisfacente nell'originarietà causale relativa: essa si limita a identificare il penultimo anello della catena causale come «origine» dell'azione, ignorando che quel penultimo anello è esso stesso determinato da anelli precedenti. Se l'originarietà causale deve essere un'originarietà nel senso in cui il georgiano la grammaticalizza — l'evento emana dall'agente come suo punto di partenza irriducibile — allora il determinismo la esclude effettivamente.

La concessione compatibilista su questo punto è dunque significativa: in un mondo deterministico, nessun agente è nel senso forte della parola il punto di partenza irriducibile della propria azione. La catena causale risale sempre al di là di lui.

Ma questa concessione ha un'importanza pratica limitata, per una ragione che l'analisi multidimensionale illumina con precisione. L'originarietà causale nel senso forte che il determinismo esclude non è la proprietà dell'agentività su cui si fondano le pratiche ordinarie di responsabilità. Quando si attribuisce responsabilità a un agente, non si chiede se la catena causale dei propri stati mentali risalga al di là della propria nascita: si chiede se l'azione proviene dall'agente nel senso che essa esprime la sua struttura psicologica, le sue intenzioni, i suoi valori — indipendentemente da come quella struttura, quelle intenzioni e quei valori siano stati formati. La responsabilità ordinaria si fonda sulle altre tre dimensioni, non sull'originarietà causale assoluta. Questo è il punto decisivo che permette di concedere l'argomento incompatibilista sull'originarietà senza che la concessione si propaghi alle altre dimensioni.

8.2 Intenzionalità e determinismo: il compatibilismo ha ragione

La seconda domanda — se l'intenzionalità sia compatibile con il determinismo — ha una risposta più netta, e il

compatibilismo ha su questo punto ragioni molto forti che non sono state adeguatamente valorizzate perché il dibattito non ha distinto questa domanda dalla precedente.

L'intenzionalità nel senso filosoficamente rilevante — agire per ragioni, sulla base di deliberazioni, in modo coerente con la struttura della propria volontà — non solo è compatibile con il determinismo: è realizzabile soltanto in un sistema causale. Agire per ragioni significa che le ragioni sono la causa dell'azione — che vi è un nesso causale tra il riconoscimento delle ragioni e la produzione dell'azione. Se le ragioni non causassero le azioni in qualche modo, il «agire per ragioni» sarebbe una fantasia: le azioni verrebbero prodotte da qualcos'altro, e il loro accordo con le ragioni sarebbe accidentale.

L'incompatibilista risponde che anche l'intenzionalità deterministica è insufficiente perché la catena causale che produce l'intenzione risale al di là dell'agente. Ma questa risposta confonde nuovamente la prima dimensione (originarietà causale) con la seconda (intenzionalità). L'intenzionalità non richiede che l'agente sia l'originatore causale assoluto dei propri stati mentali: richiede che i propri stati mentali — intenzioni, credenze, desideri — siano connessi alla propria azione nel modo causalmente appropriato, e che questa connessione costituisca il tipo di nesso che giustifica la descrizione dell'azione come intenzionale. Il fatto che gli stati mentali dell'agente siano a loro volta causalmente determinati da antecedenti non rompe il nesso tra quelli stati e l'azione: lo presuppone.

L'argomento della «manipolazione» — nella forma più raffinata di Pereboom — cerca di mostrare che l'intenzionalità deterministica è equivalente all'intenzionalità del soggetto manipolato da un neuroscienziato, e dunque che entrambe sono insufficienti per la responsabilità. Ma questo argomento confonde le dimensioni in modo diagnostico: il neuroscienziato che manipola il cervello del soggetto altera simultaneamente l'originarietà causale (l'azione origina dall'intervento del neuroscienziato), l'intenzionalità (le intenzioni sono impiantate dall'esterno, non emerse dalla struttura psicologica dell'agente) e il controllo (il soggetto non ha controllo sull'azione nel momento in cui viene compiuta). Il soggetto nel mondo deterministico ordinario ha la sua intenzionalità intatta — le sue intenzioni emergono dalla sua struttura psicologica, non da un intervento esterno — anche se quella struttura è causalmente determinata da antecedenti. L'analogia con il

manipolato neuroscientificamente è falsa perché non rispetta la differenza tra le dimensioni.

8.3 Controllo e determinismo: il compatibilismo condizionale è sufficiente

La terza domanda — se il controllo dell'agente sulla propria azione sia compatibile con il determinismo — è quella su cui il dibattito compatibilista ha prodotto le analisi più elaborate, e su cui la risposta compatibilistica è probabilmente la più convincente.

Il tipo di controllo rilevante per la responsabilità morale non è il controllo assoluto — la capacità di agire in modo completamente indipendente da qualsiasi antecedente causale — ma il **controllo condizionale** o la **responsività alle ragioni**: la capacità di agire diversamente se si avessero ragioni diverse, se le circostanze fossero diverse, se si disponesse di informazioni diverse. Questo tipo di controllo — descritto con precisione da John Martin Fischer e Mark Ravizza nella nozione di *reasons-responsive mechanism* — è pienamente compatibile con il determinismo: un sistema causalmente determinato può essere tale da produrre comportamenti diversi in risposta a inputs diversi, e questa responsività causale è esattamente il tipo di controllo che importa per la responsabilità.

La condizione di Fischer e Ravizza — che il meccanismo che produce l'azione sia il meccanismo proprio dell'agente e che sia responsivo alle ragioni in modo adeguato — introduce implicitamente una distinzione tra le dimensioni dell'agentività che la nostra analisi esplicita: il meccanismo deve essere «proprio» dell'agente (una forma di originarietà causale relativa), deve produrre l'azione in risposta alle ragioni dell'agente (una forma di intenzionalità), deve essere modulabile in risposta a ragioni diverse (il controllo condizionale), e deve essere tale da rendere l'agente il referente appropriato delle pratiche di responsabilità (la responsabilità attribuibile). Fischer e Ravizza trattano queste condizioni come aspetti di un'unica condizione complessiva piuttosto che come dimensioni distinte, ma la struttura dell'analisi è implicitamente multidimensionale.

La critica incompatibilista al controllo condizionale — nella forma di van Inwagen — è che essa non cattura la «vera» possibilità di fare diversamente: nel mondo deterministico, a parità di tutti gli antecedenti, l'agente non avrebbe potuto fare

diversamente. Ma questa critica richiede la possibilità di fare diversamente in senso «assoluto» — tenendo fissi tutti gli antecedenti — che è sia incoerente con il determinismo sia irrilevante per le pratiche ordinarie di responsabilità. Le nostre pratiche di responsabilità non richiedono mai che l'agente avrebbe potuto fare diversamente tenendo fissi tutti gli antecedenti: richiedono che avrebbe potuto fare diversamente se le circostanze fossero state diverse, se avesse avuto ragioni diverse, se avesse ricevuto informazioni diverse. Questo è il controllo condizionale, ed esso è pienamente compatibile con il determinismo.

8.4 Responsabilità attribuibile e determinismo: una distinzione decisiva

La quarta domanda — se la responsabilità attribuibile sia compatibile con il determinismo — è quella su cui le tre dimensioni precedenti convergono, e la cui analisi produce il risultato più interessante dell'intera argomentazione.

La responsabilità attribuibile — essere il referente appropriato delle pratiche di lode, biasimo, sanzione, ricompensa, e delle risposte emotive strawsoniane — non dipende dall'originarietà causale assoluta (che il determinismo esclude), né dall'intenzionalità deterministica (che il determinismo ammette), né dal controllo condizionale (che il determinismo ammette): essa dipende da una configurazione delle altre tre dimensioni che varia a seconda del tipo di responsabilità in questione.

La responsabilità penale richiede tipicamente intenzionalità significativa e controllo significativo, ma non richiede originarietà causale assoluta: il mandante di un crimine è responsabile penalmente anche se non ha «originato» causalmente l'atto nel senso georgiano. La responsabilità civile per negligenza richiede originarietà causale in senso relativo (la condotta dell'agente ha causato il danno) ma non richiede intenzionalità: si è civilmente responsabili per danni causati per negligenza anche senza intenzione di causarli. La responsabilità morale nelle relazioni interpersonali — quella più direttamente rilevante per gli atteggiamenti reattivi strawsoniani — dipende fortemente dall'intenzionalità e dalla qualità della volontà manifestata, ma è modulata anche dall'originarietà relativa e dal controllo.

In tutte queste forme di responsabilità, la condizione di compatibilità con il determinismo è soddisfatta: il determinismo esclude l'originarietà causale assoluta, ma lascia intatte l'intenzionalità deterministica, il controllo condizionale, e di conseguenza la responsabilità attribuibile che si fonda su queste. Il dibattito compatibilismo/incompatibilismo — nella misura in cui riguarda la responsabilità attribuibile — è già risolto dal compatibilismo, ma la risoluzione non è stata vista con chiarezza perché il dibattito non ha separato la responsabilità attribuibile dalle altre dimensioni.

8.5 Pereboom riesaminato: una risposta sistematica

Derk Pereboom rappresenta la posizione incompatibilista più raffinata contemporanea, e il suo «four-case argument» merita una risposta sistematica che la versione precedente di questo testo non ha sviluppato in modo sufficiente.

Il four-case argument di Pereboom costruisce una serie di quattro scenari in cui l'agente compie lo stesso atto — uccide qualcuno — ma le circostanze di produzione dell'atto variano dal più al meno intuitivamente «manipolato». Nel Caso 1, un neuroscienziato manipola direttamente i processi mentali dell'agente per produrre l'atto. Nel Caso 2, la manipolazione avviene durante l'infanzia attraverso un programma di condizionamento. Nel Caso 3, la manipolazione avviene attraverso un processo di educazione da parte di persone che non sapevano di essere strumenti di un progetto manipolativo. Nel Caso 4, l'agente vive in un mondo deterministico in cui la formazione dei suoi stati mentali segue le leggi causali ordinarie. Pereboom argomenta che non vi è una differenza moralmente rilevante tra il Caso 1 e il Caso 4: se nel Caso 1 l'agente non è moralmente responsabile, non lo è neppure nel Caso 4.

La critica che questo testo oppone all'argomento di Pereboom è precisa e va oltre la risposta disponibile nel quadro binario. Il four-case argument funziona solo se si assume che la differenza moralmente rilevante tra i casi sia una differenza di grado lungo una dimensione unica — la «libertà» come grandezza scalare — in modo che la differenza tra il Caso 1 e il Caso 4 sia solo quantitativa. Se invece si adotta il quadro multidimensionale, si vede che i casi differiscono qualitativamente rispetto a dimensioni distinte, non solo quantitativamente rispetto a una dimensione unica.

Nel Caso 1 (manipolazione diretta del cervello), tutte e quattro le dimensioni sono compromesse simultaneamente: l'originarietà causale (l'azione origina dall'intervento del neuroscienziato), l'intenzionalità (le intenzioni sono impiantate dall'esterno), il controllo (il soggetto non ha controllo nel senso rilevante), la responsabilità attribuibile (il soggetto non è il referente appropriato del biasimo). Nel Caso 4 (mondo deterministico ordinario), solo la prima dimensione è compromessa nel senso profondo: l'originarietà causale assoluta è esclusa dal determinismo. Le altre tre dimensioni — intenzionalità, controllo condizionale, responsabilità attribuibile — rimangono intatte.

Pereboom non può rispondere che anche nel Caso 4 tutte e quattro le dimensioni sono compromesse senza ricorrere all'assunzione che il determinismo comprometta l'intenzionalità e il controllo, e questa assunzione è esattamente ciò che il compatibilista contesta. L'argomento per analogia di Pereboom funziona solo se le quattro dimensioni sono una sola: se si separa la prima (originarietà causale assoluta, compromessa in tutti i casi) dalle altre tre (intatte nel Caso 4), la serie di casi non è più analoga nel modo che l'argomento richiede.

III. Verso una teoria differenziata della responsabilità morale

9.1 Il sistema positivo: responsabilità come configurazione multidimensionale

L'analisi fin qui condotta permette di abbozzare le linee di una teoria positiva della responsabilità morale che tenga conto della multidimensionalità dell'agentività. Questa teoria non pretende di essere completa: essa è necessariamente schematica in questa sede, e richiede sviluppo in ricerche successive. Ma può indicare la direzione in cui il programma di ricerca suggerito dall'analisi grammaticale dovrebbe procedere.

La proposta fondamentale è che la responsabilità morale non sia una proprietà binaria degli agenti, ma una configurazione di valori rispetto alle quattro dimensioni dell'agentività. Diversi tipi di responsabilità — penale, civile, morale interpersonale, politica, genitoriale, professionale — richiedono configurazioni diverse: essi pongono soglie diverse rispetto a ciascuna dimensione, e attribuiscono pesi diversi a ciascuna dimensione nel determinare l'entità della responsabilità.

La responsabilità penale per reati dolosi richiede alta intenzionalità e alto controllo: si è penalmente responsabili per omicidio doloso solo se si ha agito con intenzione di causare la morte (intenzionalità alta) e con piena capacità di controllo dell'azione (controllo alto). L'originarietà causale è richiesta in senso relativo (la condotta dell'imputato ha causato la morte), ma non in senso assoluto (non si richiede che la catena causale non risalga al di là dell'imputato). La responsabilità attribuibile si fonda sull'intenzionalità e sul controllo, non sull'originarietà assoluta.

La responsabilità civile per responsabilità oggettiva richiede solo originarietà causale relativa, senza intenzionalità e senza controllo in senso forte: si è oggettivamente responsabili per i danni causati dai propri animali o dalle proprie attività pericolose anche senza intenzione e senza colpa, a condizione che vi sia il nesso causale tra la propria sfera di attività e il danno. Questa forma di responsabilità non ha la stessa struttura della responsabilità penale dolosa: non richiede le stesse dimensioni, e dunque non è della stessa natura.

La responsabilità morale interpersonale — la forma di responsabilità più rilevante per gli atteggiamenti reattivi strawsoniani — ha una struttura che dipende fortemente dall'intenzionalità e dalla qualità della volontà manifestata, ma è modulata dall'originarietà relativa e dal controllo in modi che producono diverse configurazioni dell'atteggiamento reattivo appropriato. I casi fenomenologici descritti nella sezione precedente — l'offesa non intenzionale ma proveniente da una struttura psicologica dell'agente, il beneficio non intenzionalmente rivolto a noi ma prodotto dall'agente con sforzo — corrispondono a configurazioni specifiche di questo spazio multidimensionale, e richiedono atteggiamenti reattivi specifici che il sistema strawsoniano non può articolare.

9.2 Gli atteggiamenti reattivi come sistema differenziato: una proposta

Sulla base dell'analisi condotta, si può tentare di abbozzare — in modo necessariamente schematico e aperto a revisione — una tassonomia differenziata degli atteggiamenti reattivi che corrisponda alle configurazioni multidimensionali dell'agentività.

Il **risentimento** in senso pieno — nella forma descritta da Strawson — è la risposta appropriata alla configurazione in cui

l'offesa ha alta originarietà relativa, alta intenzionalità, alto controllo e alta responsabilità attribuibile. Questa è la configurazione dell'azione deliberata, consapevole, controllata, in cui l'agente è interamente il proprio atto. Il risentimento pieno è appropriato a questa configurazione.

La **critica** — che è diversa dal risentimento, più fredda e meno interpersonale — è la risposta appropriata alla configurazione in cui vi è alta originarietà relativa e alto controllo, ma intenzionalità ridotta: la persona che ha prodotto un danno non per mala volontà ma per insufficiente attenzione, cura, competenza. La critica valuta un fallimento delle capacità dell'agente, non una manifestazione di mala volontà.

Il **disappunto interpersonale** — che non è né risentimento né semplice tristezza — è la risposta appropriata alla configurazione in cui vi è originarietà relativa, ma bassa intenzionalità e basso controllo: la persona che ha prodotto un danno a causa di una struttura psicologica che non governa pienamente. Non è un agente verso cui il risentimento pieno è appropriato, perché la mala volontà non era pienamente intenzionale; ma è un agente verso cui si prova qualcosa di diverso dalla compassione neutra, perché quella struttura è «sua» in un senso che la rende rilevante per il rapporto interpersonale.

L'**indignazione strutturale** — introdotta nella sezione precedente — è la risposta appropriata alla configurazione in cui vi è alta responsabilità attribuibile e alta originarietà relativa, ma intenzionalità distribuita nel tempo (intenzioni presenti nelle singole scelte che hanno costruito la struttura) piuttosto che in un singolo atto lesivo: l'agente che ha costruito e mantenuto strutture causali dannose senza compiere nessun atto singolo intenzionalmente lesivo.

La **gratitudine qualificata** — a differenza della gratitudine piena — è la risposta appropriata al beneficio prodotto dall'agente non intenzionalmente rivolto a noi ma risultato del suo sforzo e della sua competenza: si riconosce il merito dell'agente senza la dimensione dell'orientamento verso di noi che la gratitudine piena richiede.

Questa tassonomia è preliminare e certamente incompleta. Ma essa mostra che il programma di ricerca suggerito dall'analisi grammaticale è fertile: esistono stati emotivo-valutativi distinti che corrispondono a configurazioni distinte dell'agentività, e la

filosofia ha gli strumenti per analizzarli sistematicamente una volta che disponesse del quadro concettuale adeguato.

Fenomenologia degli Atteggiamenti Reattivi Differenziati e Psicologia Morale

Il progetto fenomenologico

Perché la fenomenologia è necessaria

La tassonomia degli atteggiamenti reattivi differenziati abbozzata nella Parte Quarta della monografia è necessariamente schematica: essa identifica alcune categorie — risentimento, critica, disappunto interpersonale, indignazione strutturale, gratitudine qualificata — ma non sviluppa la fenomenologia di ciascuna. Questo sviluppo è necessario per rendere credibile la tesi che si tratti di stati genuinamente distinti e non di gradazioni di uno stesso stato, e per mostrare dove le distinzioni dimensionali dell'agentività producono differenze qualitative negli atteggiamenti reattivi piuttosto che semplici differenze di intensità.

Il metodo fenomenologico che si adopera in questa sezione è quello che Edmund Husserl chiamava «variazione eidetica»: si varia sistematicamente uno dei parametri di una situazione mentre si tengono fissi gli altri, per verificare se la variazione produce un cambiamento qualitativo nella struttura dell'esperienza piuttosto che un semplice cambiamento di grado. Nel caso degli atteggiamenti reattivi, si variano le dimensioni dell'agentività mantenendo fisso il tipo di atto e la relazione interpersonale, per verificare se le variazioni nelle dimensioni producano variazioni qualitative negli atteggiamenti reattivi appropriati.

Questo metodo non è sperimentale nel senso della psicologia empirica: produce argomenti fenomenologici, non dati statistici. La sua forza dipende dalla riconoscibilità delle descrizioni — dal fatto che il lettore riconosca gli stati descritti dalla propria esperienza — e dalla coerenza del sistema di

distinzioni proposto. Entrambe queste condizioni sono in linea di principio valutabili, anche se non attraverso la procedura standard della verifica sperimentale.

Il risentimento e le sue variazioni dimensionali

La struttura del risentimento pieno

Strawson descrive il risentimento come la risposta naturale e appropriata all'offesa intenzionale: quando qualcuno ci fa del male deliberatamente, con coscienza e controllo, manifestando con il proprio atto una mala volontà nei nostri confronti, il risentimento è la risposta che emerge naturalmente e che è moralmente appropriata. Il risentimento pieno ha una struttura che si può analizzare, usando il quadro multidimensionale, come la risposta alla configurazione in cui tutte e quattro le dimensioni sono alte: alta originarietà causale dell'offesa, alta intenzionalità, alto controllo del soggetto sul proprio atto, e di conseguenza alta responsabilità attribuibile.

La fenomenologia del risentimento pieno include: la percezione dell'atto come diretto contro di sé (il che richiede che l'atto sia intenzionale); la percezione dell'atto come proveniente dalla persona offensiva (il che richiede l'originarietà causale); il senso di essere stati trattati come qualcuno il cui benessere non merita considerazione (il che richiede che l'atto sia stato controllato e deliberato); e la direzione dell'atteggiamento verso la persona come tale piuttosto che verso la situazione o le circostanze (il che richiede che la responsabilità attribuibile sia alta). Il risentimento è dunque un atteggiamento interpersonale nel senso più pieno: è la risposta a un'altra persona che ha scelto di trattarmi in un certo modo.

La variazione della dimensione intenzionale: dalla critica al risentimento

Si mantenga fisso tutto — il tipo di atto, la relazione, il danno prodotto — e si vari la dimensione dell'intenzionalità. Considerare due situazioni parallele: nella prima, un collega di lavoro mi esclude deliberatamente da una decisione importante che mi riguarda, sapendo che avrei dovuto essere coinvolto e scegliendo di non coinvolgermi. Nella seconda, lo stesso collega prende la stessa decisione dimenticandosi che avrei dovuto essere coinvolto — non per mala fede, ma per la

distrazione e il disordine che caratterizzano il suo modo di lavorare.

La risposta appropriata nella prima situazione è il risentimento: l'atto era intenzionale, mi ha escluso deliberatamente, e il risentimento è la risposta alla sua mala volontà nei miei confronti. La risposta appropriata nella seconda situazione è qualcosa di diverso: non il risentimento pieno, perché non c'era mala volontà intenzionale, ma nemmeno la semplice tristezza neutrale, perché la sua distrazione cronica e il suo disordine sistematico sono qualcosa che gli appartiene, qualcosa che origina causalmente da lui e che avrebbe potuto — almeno in linea di principio — correggere. Questo «qualcosa di diverso» che si prova nella seconda situazione è ciò che ho chiamato **critica**: la risposta a un fallimento delle capacità o dell'attenzione dell'agente piuttosto che alla sua mala volontà.

La critica ha una fenomenologia distinta dal risentimento. Il risentimento è caldo, ha un elemento di reazione all'atto come gesto verso di sé; la critica è più fredda, ha un elemento di valutazione della persona rispetto a uno standard di cura o di competenza che non è stato soddisfatto. Il risentimento si dirige verso la persona come agente intenzionale che ha scelto di trattarmi male; la critica si dirige verso la persona come agente le cui capacità o abitudini hanno prodotto un risultato dannoso. Queste differenze fenomenologiche non sono sfumature di intensità: sono differenze qualitative nella struttura dell'atteggiamento.

La variazione della dimensione del controllo: dal risentimento al disappunto

Si mantenga ora fissa l'intenzionalità — l'atto era intenzionale — e si vari la dimensione del controllo. Una persona che mi offende deliberatamente in un momento di grave perturbazione emotiva — in piena crisi, in uno stato di disregolazione che compromette significativamente il suo controllo sulle proprie parole e sui propri atti — è in una posizione diversa dall'offensore che aveva pieno controllo su di sé. L'intenzionalità è presente (vuole offendere, non è una svista); ma il controllo sull'esecuzione è ridotto (la perturbazione emotiva limita la sua capacità di governare il proprio comportamento nel momento).

La risposta appropriata a questa situazione non è né il risentimento pieno né l'atteggiamento oggettivo. C'è qualcosa

che assomiglia al risentimento — l'atto era intenzionalmente diretto contro di sé — ma c'è anche qualcosa che attenua la risposta reattiva — il controllo ridotto riduce la piena imputabilità dell'atto. Il risultato è uno stato che ha la temperatura emotiva del risentimento ma la struttura valutativa di qualcosa di più complesso: si riconosce che l'offesa era intenzionale e si prova qualcosa in risposta alla mala volontà che esprimeva, ma si tiene anche conto che quella mala volontà era espressa in uno stato in cui il soggetto non era pienamente se stesso.

Questo stato — che si potrebbe chiamare **risentimento temperato dalla comprensione causale** — non è equivalente al risentimento attenuato da una scusa ordinaria (come l'incidente o l'ignoranza dei fatti). Le scuse ordinarie rimuovono la mala volontà: mostrano che il soggetto non voleva offendere. La perturbazione emotiva non rimuove la mala volontà — il soggetto voleva offendere in quel momento — ma mostra che il soggetto non era completamente in controllo di quel volere. La differenza fenomenologica è reale: sapere che qualcuno ci ha offeso perché non sapeva che ci avrebbe fatto male è diverso da sapere che ci ha offeso perché in quel momento la propria capacità di autoregolazione era compromessa. Nel primo caso, il risentimento scompare quasi completamente; nel secondo, esso si trasforma in qualcosa di più complesso che include sia la risposta all'offesa sia la comprensione della sua genesi.

La variazione della dimensione dell'originarietà causale: la complicità e l'intermediario

Si vari ora la dimensione dell'originarietà causale, mantenendo fissi intenzionalità e controllo. Un agente che esegue deliberatamente e con pieno controllo un atto dannoso su ordine di qualcun altro — un soldato che esegue un ordine illegale, un esecutore che porta a termine un piano che altri hanno progettato — ha alta intenzionalità e alto controllo ma originarietà causale distribuita tra lui e il soggetto che ha progettato e ordinato l'atto.

La risposta appropriata a questo agente è diversa dalla risposta all'originatore del piano. C'è qualcosa del risentimento o dell'indignazione, perché l'atto era intenzionale e controllato: il soggetto ha scelto di eseguirlo, avrebbe potuto rifiutarsi. Ma c'è anche qualcosa che distingue questa risposta da quella verso l'originatore: l'esecutore non aveva la stessa originarietà causale profonda dell'ordinante, non era il punto da cui l'idea e la

pianificazione dell'atto sono partite. Questo produce un atteggiamento che si potrebbe chiamare **indignazione qualificata dall'attribuzione parziale**: si prova indignazione per l'esecuzione consapevole dell'atto, ma si riserva qualcosa di diverso — e di più intenso — per l'originatore del piano.

La distinzione tra l'esecutore e l'originatore è al cuore di alcune delle dispute morali e giuridiche più difficili del Novecento: i processi di Norimberga, le discussioni sulla responsabilità dei soldati per crimini di guerra, il dibattito sull'obbedienza agli ordini come attenuante o come aggravante. La difficoltà di queste dispute non è solo pratica o giuridica: è filosofico-concettuale. Il sistema binario degli atteggiamenti reattivi non offre strumenti per distinguere in modo preciso tra la responsabilità dell'esecutore e quella dell'ordinante rispetto a dimensioni diverse dell'agentività, e il risultato è che le discussioni si svolgono in modo confuso, con argomenti che parlano di piani diversi senza rendersene conto.

Il problema degli atteggiamenti reattivi nel tempo

La diacronia degli atteggiamenti reattivi

Un aspetto degli atteggiamenti reattivi che la trattazione di Strawson e dei suoi successori tende a ignorare è la loro struttura temporale: gli atteggiamenti reattivi non sono stati statici che si hanno o non si hanno in un momento, ma processi dinamici che si sviluppano nel tempo e che possono cambiare in risposta a nuove informazioni, a cambiamenti nell'agente offensore, a riflessione e a elaborazione.

Il risentimento può trasformarsi in perdono quando l'offensore riconosce l'offesa, la comprende nelle sue dimensioni, e fa qualcosa per riparare. Ma il percorso dal risentimento al perdono non è uniforme: esso dipende da quale delle dimensioni dell'agentività fosse compromessa nell'atto offensivo, e da come il riconoscimento e la riparazione dell'offensore modificano quella compromissione.

Se l'offesa aveva alta intenzionalità, il riconoscimento della propria mala volontà da parte dell'offensore è necessario per avviare il percorso verso il perdono: senza riconoscimento dell'intenzionalità, il percorso non inizia, perché ciò che è stato offeso — il senso di essere stati trattati con mala volontà —

non viene toccato dalla riparazione materiale del danno. Se l'offesa aveva bassa intenzionalità ma alta originarietà causale — era il prodotto di una struttura psicologica dell'offensore, non di mala volontà nel senso stretto — il percorso verso la riconciliazione interpersonale richiede qualcosa di diverso dal riconoscimento della mala volontà: richiede la comprensione della struttura psicologica che ha prodotto l'atto, e la comunicazione di questa comprensione, e magari l'impegno a lavorare su quella struttura.

La diacronia degli atteggiamenti reattivi è dunque strutturata in modo dipendente dalle dimensioni dell'agentività che li hanno generati, e le pratiche di riconciliazione, perdono, e restaurazione delle relazioni interpersonali dipendono da questa struttura dimensionale in modo che le teorie della responsabilità non hanno ancora analizzato sistematicamente.

La giustizia riparativa come pratica multidimensionale

Il movimento della **giustizia riparativa** — che si è sviluppato a partire dagli anni Settanta come alternativa o complemento ai sistemi punitivi tradizionali — ha elaborato, nella pratica, soluzioni che riflettono implicitamente la struttura multidimensionale degli atteggiamenti reattivi, anche senza il supporto di una teoria esplicita che le giustifichi.

I programmi di giustizia riparativa — come i Victim-Offender Mediation Programs, i circoli di pacificazione ispirati alle pratiche indigene nordamericane, i processi di Truth and Reconciliation come quello sudafricano — non trattano la responsabilità come una condizione binaria che si ha o non si ha: essi strutturano un processo in cui le diverse dimensioni dell'agentività vengono esplicitamente affrontate in sequenza. Il riconoscimento dell'originarietà causale (l'offensore ha causato il danno); l'esplorazione dell'intenzionalità (perché ha agito in quel modo, cosa voleva, cosa pensava); la valutazione del controllo (in quali circostanze aveva o non aveva controllo sulla propria condotta); e infine la questione della responsabilità attribuibile nel senso prospettico (cosa l'offensore è tenuto a fare per riparare, e di cosa è capace).

Il fatto che le pratiche di giustizia riparativa si strutturino in questo modo — esplicitando dimensioni che i sistemi punitivi tradizionali tendono a collassare in una condanna binaria di colpevolezza — è un'evidenza pratica della necessità delle

distinzioni dimensionali. Non si afferma che i praticanti della giustizia riparativa abbiano consapevolmente adottato il quadro multidimensionale: si afferma che la pratica ha scoperto empiricamente la necessità di distinzioni che la teoria non aveva ancora sistematizzato.

La psicologia clinica e le dimensioni dell'agentività

Autoresponsabilità e psicoterapia

La psicologia clinica ha sviluppato, spesso in modo indipendente dalla filosofia, strumenti terapeutici che operano sulle dimensioni dell'agentività in modo differenziato. Questa convergenza pratica è ulteriore evidenza della necessità delle distinzioni dimensionali: quando sistemi di pratica sviluppati da necessità cliniche convergono con categorie derivate dall'analisi filosofica, entrambe le fonti si corroborano reciprocamente.

La terapia cognitivo-comportamentale (CBT) lavora principalmente sulla dimensione del controllo: il suo obiettivo centrale è aiutare il paziente a sviluppare capacità di controllo su processi che erano automatici, impulsivi, o guidati da schemi cognitivi disfunzionali. Le tecniche di CBT — la ristrutturazione cognitiva, il training di skills, il problem-solving strutturato — sono strumenti per aumentare il controllo sull'esecuzione dei propri atti. Non operano primariamente sull'intenzionalità (il paziente sa spesso già cosa dovrebbe fare) né sull'originarietà causale (la CBT non si preoccupa molto della storia causale dei problemi del paziente), ma specificamente sul controllo.

La terapia psicotica lavora principalmente sulla dimensione dell'intenzionalità: il suo obiettivo centrale è rendere consapevoli intenzioni che erano latenti, dinamiche motivazionali che guidavano il comportamento senza essere riconosciute come intenzioni. Il processo di elaborazione psicoanalitica — il lavoro sull'interpretazione dei sogni, delle fantasie, dei transfert — è un processo di portare alla coscienza intenzioni che operavano al di sotto della soglia della deliberazione esplicita. Non opera primariamente sul controllo (portare alla coscienza un'intenzione non automaticamente aumenta il controllo sulla sua traduzione in atto) né sull'originarietà causale nel senso profondo, ma specificamente sull'intenzionalità.

Le terapie orientate alla narrativa — la narrative therapy di White e Epston, la logoterapia di Frankl, alcune forme di terapia esistenziale — lavorano principalmente sulla dimensione della responsabilità attribuibile: esse aiutano il paziente a ridefinire la propria storia in modo da ricollocare la responsabilità in modo più adeguato, riconoscendo dove si ha agentività reale e dove si è stati vittime di circostanze. Il processo terapeutico in questi approcci non è quello di aumentare il controllo o di rendere coscienti le intenzioni latenti, ma di rinegoziare la struttura narrativa dell'attribuzione di responsabilità.

L'ACT — Acceptance and Commitment Therapy — lavora in modo più complesso su più dimensioni simultaneamente: l'accettazione degli stati psicologici senza identificarsi con essi lavora sulla dimensione dell'intenzionalità (distinguendo tra stati che si «hanno» e stati che si «è»); il lavoro sui valori e sull'impegno lavora sulla dimensione del controllo prospettico; la defusione cognitiva lavora sulla dimensione dell'originarietà causale (riconoscendo che certi pensieri automatici non «origina»no dall'agente nel senso che li ha scelti).

Il problema della vergogna e del senso di colpa come evidenza delle dimensioni

La distinzione tra vergogna e senso di colpa è uno dei temi più studiati nella psicologia morale clinica, e offre un'evidenza indipendente della necessità delle distinzioni dimensionali.

La vergogna — nel senso clinico — è un'emozione che riguarda il sé come totalità: quando si prova vergogna, si ha la percezione di essere fondamentalmente difettosi, inadeguati, non meritevoli. Il senso di colpa, al contrario, riguarda un atto specifico: quando si prova senso di colpa, si ha la percezione di aver fatto qualcosa di sbagliato senza che questo implichi necessariamente un giudizio negativo sulla propria persona come totalità.

June Price Tangney, in una serie di studi sistematici sulla distinzione vergogna/colpa, ha mostrato che queste due emozioni hanno strutture diverse, producono risposte comportamentali diverse, e hanno diverse implicazioni per la salute mentale e per la condotta morale. La vergogna tende a produrre ritiro, evitamento, e talvolta esternalizzazione difensiva dell'aggressività; il senso di colpa tende a produrre riparazione, impegno a non ripetere, e ricerca di riconciliazione.

Il senso di colpa ha generalmente effetti più positivi sulla condotta morale della vergogna.

Questa distinzione clinica tra vergogna e colpa corrisponde, nel quadro multidimensionale, alla distinzione tra una risposta che riguarda l'originarietà causale nella sua forma più profonda — la vergogna colpisce il sé come sorgente causale di certi atti — e una risposta che riguarda atti specifici nella loro intenzionalità e controllabilità — la colpa riguarda ciò che si ha fatto piuttosto che ciò che si è. La vergogna è la risposta a una configurazione in cui si percepisce come compromessa la prima dimensione nella sua forma più profonda: si sente di essere, come persona, una sorgente di danno. Il senso di colpa è la risposta a una configurazione in cui si riconosce che un atto specifico aveva alta intenzionalità e controllo, e che dunque la responsabilità è pienamente attribuibile — ma senza che questo implichi un giudizio sulla propria struttura causale come persona.

La distinzione è clinicamente rilevante perché la vergogna eccessiva — la tendenza a rispondere agli errori e alle offese con vergogna piuttosto che con senso di colpa — è associata a peggior outcome di salute mentale, a maggiore difficoltà nelle relazioni interpersonali, e a minore capacità di riparazione morale. Le terapie che cercano di spostare i pazienti dalla vergogna al senso di colpa — come alcuni approcci della terapia focalizzata sulla compassione — lavorano implicitamente sulla distinzione dimensionale: aiutano il paziente a separare la valutazione dell'atto specifico (alta intenzionalità e responsabilità attribuibile per quell'atto) dalla valutazione del sé come sorgente causale (che non dovrebbe essere compromessa da ogni singolo atto sbagliato).

L'empatia come risposta multidimensionale

La struttura multidimensionale dell'empatia morale

L'empatia — la capacità di rispondere agli stati emotivi degli altri in modo appropriato — è una delle componenti più discusse della psicologia morale, e ha generato una letteratura enorme sia in psicologia che in filosofia. Ma la maggior parte di questa letteratura tratta l'empatia come una capacità unitaria — si è empatici o non lo si è — senza distinguere le diverse

componenti che la compongono e le diverse dimensioni dell'agentività a cui essa risponde.

Paul Bloom, in *Against Empathy* (2016), ha sollevato una critica provocatoria all'empatia come fondamento della moralità, sostenendo che l'empatia — intesa come la capacità di sentire ciò che l'altro sente — sia in realtà un meccanismo cognitivo distorto che favorisce coloro che ci sono vicini a scapito di quelli lontani, e che sia dunque un pessimo fondamento per decisioni morali imparziali. Bloom propone di sostituirla con una «preoccupazione razionale» per il benessere altrui, che non dipende dalla capacità di sentire ciò che l'altro sente ma dalla comprensione cognitiva delle sue esigenze.

La critica di Bloom è parzialmente giusta ma confonde sotto il termine «empatia» meccanismi distinti. C'è la **simulazione emotiva** — il sentire ciò che l'altro sente, il meccanismo descritto da Bloom come distorto e partigiano — e c'è la **comprensione empatica** — la capacità di capire le motivazioni, le intenzioni, le prospettive dell'altro anche senza dividerne l'emozione. Queste due capacità corrispondono, nel quadro dimensionale, a due diversi livelli di risposta all'agentività altrui: la simulazione emotiva risponde agli stati psicologici dell'altro come locus di certi stati (la dimensione dativa, per riprendere la terminologia georgiana); la comprensione empatica risponde all'altro come agente con intenzioni e originarietà causale.

Una psicologia morale adeguata ha bisogno di entrambe le forme di empatia, e ha bisogno di distinguerle: la simulazione emotiva è preziosa per le relazioni di prossimità e per la comprensione immediata degli stati emotivi; la comprensione empatica è necessaria per la valutazione morale degli atti degli altri, per la risposta alle loro intenzioni e alla loro agentività. Le distorsioni che Bloom attribuisce all'empatia riguardano primariamente la simulazione emotiva; la comprensione empatica non ha necessariamente le stesse distorsioni.

Implicazioni per la Teoria Giuridica

I. Il diritto come sistema di attribuzione dell'agentività

11.1 Il diritto come filosofia dell'azione istituzionalizzata

Il diritto penale, il diritto civile e il diritto amministrativo sono, nella loro struttura più profonda, sistemi istituzionalizzati di attribuzione dell'agentività: essi devono decidere, per ogni danno o illecito che si produce nel tessuto sociale, chi è l'agente rilevante, in quale senso lo è, e quali conseguenze questa attribuzione produce. Questa funzione li rende il laboratorio pratico più ricco per verificare se la quadripartizione dimensionale dell'agentività proposta in questo testo abbia tenuta al di fuori del dominio puramente filosofico.

Il diritto non ha il lusso dell'astrazione indefinita: esso deve produrre decisioni, e le decisioni devono essere motivate con ragioni che resistano all'esame pubblico e al controllo giurisdizionale. Quando la struttura concettuale di cui il diritto dispone è inadeguata alla complessità del caso, questa inadeguatezza si manifesta in modo tangibile: nelle incongruenze della giurisprudenza, nelle motivazioni che scivolano tra criteri diversi senza riconoscerlo, nelle dottrine eccezionali costruite ad hoc per aggirare le aporie del sistema ordinario. L'analisi delle aporie del diritto è dunque un modo di verificare se le dimensioni dell'agentività identificate in questo testo siano effettivamente distinte nella pratica, non solo nella teoria.

La tesi che si argomenta in questa sezione è che il diritto occidentale — sviluppato in lingue indoeuropee e strutturato dalla loro grammatica binaria dell'agentività — abbia prodotto sistemi di attribuzione della responsabilità che funzionano bene per i casi paradigmatici (azione intenzionale, controllata, originata dall'agente, con danno direttamente causato) ma producono risultati insoddisfacenti o incoerenti per tutti i casi in cui le quattro dimensioni si disaccoppiano. Questi casi non sono eccezioni marginali: essi sono al centro di alcune delle questioni più urgenti del diritto contemporaneo — la responsabilità delle persone giuridiche, la responsabilità per danni ambientali a lungo termine, la responsabilità per i crimini commessi in contesti di obbedienza istituzionale, la responsabilità dei sistemi di intelligenza artificiale.

11.2 La struttura della mens rea: una tassonomia implicita e le sue lacune

Il diritto penale anglosassone e, in misura analoga, le tradizioni di civil law europee articolano l'elemento soggettivo del reato — la *mens rea* — attraverso una serie di categorie che rappresentano, storicamente, un tentativo di differenziare i gradi di agentività rilevante per la punibilità. Le categorie principali — dolo intenzionale, dolo eventuale, colpa cosciente, colpa incosciente, responsabilità oggettiva — sono ordinabili lungo quello che la tradizione giuridica tratta come un continuum dal più al meno «volontario». Ma questa struttura, esaminata alla luce della quadripartizione dimensionale, rivela una serie di incongruenze sistematiche che non sono accidentali.

Il **dolo intenzionale** corrisponde alla configurazione in cui tutte e quattro le dimensioni sono pienamente soddisfatte: l'agente è l'originatore causale dell'evento, lo ha inteso, lo ha controllato nell'esecuzione, e ne è pienamente responsabile. Questa è la configurazione paradigmatica, e il diritto la tratta con la massima severità punitiva.

Il **dolo eventuale** — nella sistematica italiana, l'equivalente di ciò che il diritto inglese chiama *recklessness* — introduce una frattura nella struttura paradigmatica: l'agente non ha inteso l'evento come scopo della propria azione, ma ha previsto la possibilità che esso si verificasse come conseguenza della propria condotta e ha accettato questo rischio. Qui l'intenzionalità rispetto all'evento è ridotta (l'evento non era lo scopo) ma non azzerata (il rischio era accettato); l'originarietà causale rimane alta; il controllo era presente ma non esercitato per evitare il rischio. La struttura dimensionale di questa configurazione è significativamente diversa dal dolo intenzionale, eppure il diritto la tratta come variante del dolo — e dunque punisce con la stessa cornice di pena — producendo tensioni sistematiche nella giurisprudenza che tenta di distinguere il dolo eventuale dalla colpa cosciente senza disporre di criteri concettuali adeguati.

La **colpa cosciente** — la previsione del rischio senza l'accettazione del rischio verificarsi — differisce dal dolo eventuale esattamente rispetto alla dimensione dell'intenzionalità parziale: nel dolo eventuale vi è accettazione (una forma degradata di intenzione); nella colpa cosciente vi è previsione senza accettazione. Ma il confine tra accettare un rischio e prevederlo senza accettarlo è notoriamente difficile da tracciare nella giurisprudenza concreta, e le difficoltà sono sistematiche piuttosto che accidentali: esse derivano dal fatto che il diritto cerca di distinguere due configurazioni

dimensionalmente diverse — diverse rispetto all'intenzionalità — attraverso un test psicologico introspettivo (l'agente «accettava» il rischio?) che non ha una risposta univoca nella maggioranza dei casi concreti.

La **colpa incosciente** — la mancata previsione di un rischio che l'agente avrebbe dovuto prevedere — introduce una variabile ulteriore che la struttura dimensionale permette di articolare con precisione: qui l'intenzionalità è assente non solo rispetto all'evento ma rispetto al rischio; il controllo era potenzialmente disponibile ma non esercitato per via di una carenza di attenzione o di competenza; l'originarietà causale rimane in senso relativo. La configurazione dimensionale è diversa da tutte le precedenti, e il diritto la tratta come categoria autonoma. Ma la struttura concettuale con cui la giustifica — la colpa come deviazione dallo standard dell'agente ragionevole — è costruita con strumenti diversi rispetto a quelli usati per il dolo, senza riconoscere che essa e il dolo differiscono lungo dimensioni identificabili.

La **responsabilità oggettiva** — l'attribuzione di responsabilità senza requisito di colpa o dolo, sulla base del solo nesso causale tra condotta e danno — è la categoria che produce le maggiori difficoltà concettuali nel pensiero giuridico tradizionale, perché essa sembra attribuire responsabilità in assenza di agentività soggettiva. Dal punto di vista della quadripartizione dimensionale, la responsabilità oggettiva è spiegabile in modo più soddisfacente: essa richiede solo la dimensione dell'originarietà causale relativa (la condotta dell'agente ha causato il danno), prescindendo dall'intenzionalità e dal controllo. Non è «responsabilità senza agentività»: è responsabilità fondata su una singola dimensione dell'agentività. Questo la rende concettualmente più trasparente, e permette di vedere perché essa è appropriata in certi contesti (attività pericolose, responsabilità del produttore, responsabilità ambientale) e non in altri: è appropriata quando la dimensione causale è sufficiente per fondare la pratica di attribuzione rilevante, anche in assenza delle dimensioni dell'intenzionalità e del controllo.

11.3 Il problema della causalità giuridica: l'originarietà distribuita

Tra i problemi più ostici della teoria del diritto penale e civile vi è quello della **causalità**: determinare quando la condotta di un agente è la «causa» del danno ai fini giuridici, in modo da

fondare il nesso di causalità necessario per l'imputazione. La difficoltà non è puramente fattuale — non è solo la questione di stabilire cosa è successo — ma concettuale: i criteri di causalità giuridicamente rilevante divergono sistematicamente dai criteri della causalità scientifica, e questa divergenza è rivelatrice.

La teoria della *condicio sine qua non* — secondo cui la condotta dell'agente è causa del danno se, eliminata mentalmente la condotta, il danno non si sarebbe verificato — produce risultati assurdi nei casi di causalità alternativa ipotetica (se io non avessi ucciso la vittima, l'avrebbe uccisa qualcun altro) e di causalità cumulativa (due agenti indipendenti producono insieme un danno che nessuno dei due avrebbe prodotto da solo). La teoria della causalità adeguata — secondo cui è causa giuridicamente rilevante solo la condotta che era statisticamente adeguata a produrre l'evento del tipo verificatosi — produce risultati difendibili nei casi ordinari ma si scontra con difficoltà nei casi di decorso causale atipico. La teoria dell'imputazione obiettiva — sviluppata nella dogmatica tedesca e influente nel diritto comparato — introduce il requisito della «creazione di un rischio non consentito» come condizione della causalità giuridicamente rilevante, avvicinandosi a una struttura multidimensionale ma senza formalizzarla esplicitamente.

L'analisi dimensionale permette di vedere queste difficoltà sotto una luce diversa. Il problema della causalità giuridica è in parte il problema di come la dimensione dell'originarietà causale si distribuisce tra più agenti, tra agenti e strutture, tra azioni e omissioni, in modo da fondare la responsabilità attribuibile in ciascuno di questi casi. Non vi è un concetto unitario di «causazione giuridicamente rilevante» che si applica uniformemente a tutti questi casi: vi sono configurazioni diverse dell'originarietà causale che corrispondono a forme diverse di responsabilità attribuibile.

Il caso della **causalità omissiva** illustra questa tesi con particolare chiarezza. Chi omette di prestare soccorso a una persona in pericolo — quando ne aveva l'obbligo — è «causa» della morte di quella persona? La risposta della logica causale ordinaria è negativa: l'omissione non è un evento positivo che produce causalmente l'esito. Eppure il diritto attribuisce responsabilità all'omittente nei casi in cui vi era obbligo di intervento. Dal punto di vista dimensionale, l'omittente ha originarietà causale in senso negativo (la sua non-azione ha lasciato spazio al processo causale che ha prodotto il danno), aveva controllo potenziale non esercitato, ma ha intenzionalità

solo rispetto all'omissione, non rispetto all'esito. Questa configurazione dimensionale è diversa da quella dell'azione causale diretta, e il diritto la tratta come categoria distinta con requisiti propri. Ma la struttura concettuale con cui lo giustifica — la distinzione tra causazione attiva e causazione per omissione — non è sufficientemente articolata da spiegare perché certi obblighi di intervento generino responsabilità e altri no.

1.4 Merleau-Ponty e l'agentività incorporata: un anticipo filosofico

Prima di procedere all'analisi del sistema georgiano, è utile segnalare che la tradizione filosofica continentale ha prodotto, in modo indipendente dalla tipologia linguistica, intuizioni sull'agentività che convergono con il quadro multidimensionale che questo testo sviluppa. Maurice Merleau-Ponty, nella *Fenomenologia della percezione* (1945), ha elaborato una critica della concezione soggetto-centrica dell'azione che anticipa in molti punti le insufficienze del modello binario indoeuropeo. Per Merleau-Ponty, l'azione non è l'effetto di una volontà che comanda un corpo passivo: è una struttura di risposta in cui il corpo-soggetto si orienta verso il mondo attraverso schemi motori acquisiti che non sono pienamente intenzionali nel senso cartesiano, ma non sono nemmeno puramente meccanici. Il concetto di «intenzionalità motoria» — la direzionalità del corpo verso il suo ambiente senza la mediazione esplicita della rappresentazione cosciente — cattura esattamente quella zona dell'agentività che le lingue indoeuropee confondono nel soggetto nominativo.

La ragione per cui questo anticipo fenomenologico non ha trovato sbocco nella filosofia analitica dell'azione è in parte istituzionale — le due tradizioni hanno comunicato poco — e in parte grammaticale. Merleau-Ponty scrive in francese, una lingua con lo stesso schema nominativo-accusativo del latino e del greco che questa trattazione esamina. Per esprimere l'agentività intermedia che la voce media greca o il dativo georgiano grammaticalizzano, egli deve costruire costruzioni perifrasi complesse e spesso oscure — «il corpo che si muove verso», «l'intenzionalità operante», «l'arco intenzionale» — che cercano di dire ciò che la grammatica francese non dice automaticamente. Il risultato è una prosa di grande potenza filosofica ma di difficile esegesi, la cui oscurità è in parte intrinseca alla difficoltà di esprimere in una grammatica binaria strutture che richiederebbero una grammatica graduale. Il

sistema georgiano non avrebbe reso Merleau-Ponty superfluo; avrebbe reso la sua prosa più perspicua.

Questa convergenza tra fenomenologia continentale e tipologia linguistica non è accidentale. Entrambe identificano — attraverso strumenti e metodi molto diversi — la stessa insufficienza del modello binario dell'agentività: la sua incapacità di rendere conto delle zone intermedie tra l'azione intenzionale piena e la passività totale, zone che costituiscono, come l'esperienza ordinaria conferma, la stragrande maggioranza degli eventi che co-determiniamo. Questa convergenza è anche un argomento indiretto a favore della solidità filosofica del quadro multidimensionale: esso non è un'importazione esotica dalla tipologia linguistica, ma articola sistematicamente una struttura che pensatori diversi, in tradizioni diverse, hanno riconosciuto come necessaria attraverso approcci indipendenti.

11.4 La responsabilità delle persone giuridiche: agentività senza soggetto nominativo

La responsabilità penale delle persone giuridiche — le organizzazioni, le società, le istituzioni — è uno dei campi in cui il deficit concettuale della struttura binaria dell'agentività si manifesta con maggiore evidenza. Le tradizioni giuridiche di common law e civil law hanno sviluppato approcci diversi a questo problema, e le loro differenze riflettono approcci diversi alla stessa difficoltà fondamentale: come si attribuisce responsabilità a un'entità collettiva che non ha soggettività psicologica nel senso in cui ce l'ha un individuo?

La difficoltà concettuale nasce dalla struttura grammaticale: il soggetto nominativo della proposizione è prototipicamente un individuo, e il soggetto della responsabilità penale è storicamente costruito sul modello dell'individuo con stati mentali (intenzioni, volontà, deliberazioni). Una persona giuridica non ha stati mentali in questo senso: non «vuole» nel senso in cui vuole un essere umano, non «delibera» nel senso in cui delibera un individuo. La tradizione giuridica ha cercato di risolvere questo problema attraverso la finzione della volontà collettiva — trattando la volontà degli organi della persona giuridica come volontà della persona giuridica stessa — ma questa finzione produce risultati insoddisfacenti nei casi in cui la cultura organizzativa produce comportamenti illeciti senza che nessun individuo specifico abbia «voluto» quell'esito.

Dal punto di vista della quadripartizione dimensionale, il problema della responsabilità delle persone giuridiche è il problema di come le dimensioni dell'agentività si ridistribuiscono nelle strutture collettive. Una persona giuridica può avere alta originarietà causale relativa (le sue attività hanno causato il danno), strutture di controllo istituzionalizzato (procedure, politiche, incentivi), e responsabilità attribuibile in senso istituzionale (essa è il referente appropriato di certi tipi di sanzione), senza che vi sia una corrispondente intenzionalità individuale identificabile. La configurazione dimensionale di questa forma di agentività è specificamente diversa da quella individuale, e richiede un sistema di attribuzione della responsabilità che non sia semplicemente la proiezione del modello individuale sul collettivo.

La legislazione italiana sul tema — il Decreto Legislativo 231/2001 — ha introdotto un sistema di responsabilità amministrativa degli enti per i reati commessi nell'interesse dell'ente da suoi organi e rappresentanti. Questo sistema è ibrido: attribuisce responsabilità all'ente pur riconoscendo che l'ente non ha dolo o colpa nel senso individuale, fondando la responsabilità sulla «colpa organizzativa» — la mancata adozione di modelli organizzativi idonei a prevenire il reato. Questo concetto di «colpa organizzativa» è un tentativo di articolare una forma di responsabilità collettiva che non sia riducibile alla responsabilità individuale, e che si fondi sulla dimensione del controllo istituzionale piuttosto che sull'intenzionalità individuale. Il punto interessante, dal punto di vista di questo testo, è che il legislatore ha dovuto inventare questo concetto ad hoc — la «colpa organizzativa» non aveva precedenti chiari nella dogmatica penalistica — perché il sistema concettuale ordinario non disponeva di strumenti per pensare la responsabilità collettiva in modo non riduttivo rispetto al modello individuale.

11.5 I crimini commessi per ordine: la responsabilità dell'esecutore

Il problema della responsabilità penale di chi compie atti criminosi in esecuzione di ordini superiori — «executing superior orders» nella terminologia del diritto internazionale — è uno dei più tormentati della storia del diritto, e la sua tormentata storia riflette l'inadeguatezza degli strumenti concettuali disponibili per trattarlo.

La risposta del diritto internazionale dei diritti umani, cristallizzata nei processi di Norimberga e poi nel diritto internazionale penale contemporaneo, è che l'esecuzione di ordini superiori non costituisce una causa di giustificazione né di esclusione della colpevolezza per i crimini di guerra e contro l'umanità. Questo principio è giuridicamente consolidato, ma la sua giustificazione concettuale rimane debole nella misura in cui si confronta con la complessità effettiva delle situazioni in cui gli ordini vengono eseguiti.

Chi esegue un ordine criminoso in un sistema gerarchico coercitivo — dove la disobbedienza è punita con la morte — ha una configurazione dimensionale dell'agentività che è genuinamente complessa. L'originarietà causale dell'atto è presente in senso relativo (l'esecutore ha compiuto fisicamente l'atto) ma distribuita tra lui e i suoi superiori. L'intenzionalità rispetto all'atto singolo è presente (egli vuole compierlo per evitare la punizione) ma l'intenzionalità rispetto all'esito finale può essere assente o avversata. Il controllo sull'esecuzione è presente, ma il controllo sulla situazione che lo ha posto di fronte a quella scelta era assente. La responsabilità attribuibile è parziale: vi è qualcosa da attribuire all'esecutore, ma in una forma che è qualitativamente diversa dalla responsabilità del pianificatore o del comandante.

Il principio di Norimberga ignora questa complessità dimensionale e tratta la responsabilità dell'esecutore come binaria: o l'ordine costituisce causa di giustificazione (e allora l'esecutore è pienamente esente) oppure non la costituisce (e allora è pienamente responsabile). La dottrina successiva ha introdotto temperamenti — la costrizione come attenuante, la «duress» come causa di esclusione parziale della responsabilità — ma senza un quadro concettuale che permetta di articolare sistematicamente perché certi gradi di costrizione producono certi effetti sulla responsabilità e non altri.

Hannah Arendt, nella sua analisi del caso Eichmann, ha posto in termini filosofici la questione di questo deficit concettuale: Eichmann non era il tipo di criminale che i concetti giuridici ordinari erano stati costruiti per trattare. Non era il sadico che godeva del male commesso; non era il fanatismo ideologico che perseguitava il male come scopo. Era qualcosa di diverso e di più inquietante: un burocrate che eseguiva ordini senza «pensare» — senza esercitare il giudizio morale che avrebbe potuto interrompere l'esecuzione. La «banalità del male» di Arendt è, tradotta nel linguaggio di questo testo, la descrizione di una configurazione dell'agentività in cui l'intenzionalità rispetto ai

singoli atti era presente (Eichmann voleva compiere ciò che compiva) ma l'intenzione rispetto agli effetti era sospesa o assente, e il controllo era presente nell'esecuzione ma non nel governo della propria situazione complessiva. Questa configurazione non ha nome nel vocabolario giuridico ordinario, e il processo Eichmann — come Arendt ha mostrato con amarezza — ha faticato a trovare le categorie giuridiche adeguate per trattarla.

11.6 Responsabilità per danni ambientali: originarietà distribuita nel tempo

Il diritto ambientale contemporaneo ha dovuto fare i conti con una forma di agentività che la struttura concettuale tradizionale non era attrezzata a gestire: la responsabilità per danni che si producono a lungo termine come risultato di azioni che al momento della loro esecuzione erano legalmente permesse e prive di effetti dannosi immediatamente percepibili.

Il caso dell'inquinamento industriale è paradigmatico. Un'impresa che ha scaricato sostanze chimiche in un corso d'acqua nel 1970 — quando questo era consentito dalla legislazione vigente e gli effetti a lungo termine sulla falda freatica non erano noti — produce danni che si manifestano trent'anni dopo. Chi è l'agente responsabile? In che senso lo è? Rispetto a quali dimensioni dell'agentività si fonda la responsabilità?

Il diritto vigente risponde in modo diverso in diversi ordinamenti, e le divergenze non sono superficiali ma riflettono concezioni diverse della struttura della responsabilità. Il diritto americano della responsabilità ambientale — in particolare il CERCLA (Comprehensive Environmental Response, Compensation, and Liability Act) — adotta un principio di responsabilità retroattiva e oggettiva: chi ha contribuito alla contaminazione di un sito è responsabile dei costi di bonifica anche se la condotta era lecita al momento in cui è stata tenuta. Questa soluzione fondata sull'originarietà causale indipendentemente dall'intenzionalità e dal controllo è giustificabile come politica ambientale — essa produce incentivi efficaci alla cautela — ma è concettualmente insoddisfacente come teoria della responsabilità: essa non distingue tra l'impresa che ha agito con piena consapevolezza del rischio e l'impresa che ha agito in buona fede secondo gli standard del tempo.

La distinzione rilevante, nella prospettiva dimensionale, è tra l'intenzionalità rispetto al danno (assente in entrambi i casi, se non si sapeva del danno) e il controllo potenziale sul rischio (che potrebbe essere diverso: un'impresa potrebbe aver avuto segnali di rischio che ha ignorato, mentre un'altra no). Il diritto ambientale che si fonda esclusivamente sull'originarietà causale non può distinguere questi due casi, e produce risultati percepiti come ingiusti da chi aveva agito in buona fede senza avere accesso alle informazioni necessarie per prevedere il danno.

11.6bis La responsabilit  algoritmica e le quattro dimensioni: una mappa

Prima di affrontare il caso dei sistemi di intelligenza artificiale in senso stretto, e utile costruire una mappa della responsabilit  algoritmica usando il quadro multidimensionale. Questa mappa permette di distinguere tra categorie di agentivit  algoritmica che il dibattito corrente tende a confondere, e di identificare dove si trovano le tensioni concettuali genuine e dove si trovano invece difficolt  artificiali prodotte dalla presupposizione di un modello unitario della responsabilit .

Il primo tipo di agentivit  algoritmica   quello dei sistemi che eseguono istruzioni determinate in modo diretto: un software di contabilit  che calcola le imposte, un sistema di controllo industriale che regola la temperatura di un reattore. In questi sistemi, la dimensione del controllo esecutivo   interamente trasferita al sistema, ma le dimensioni dell'originariet  causale e dell'intenzionalit  rimangono saldamente negli ingegneri e nei committenti che hanno specificato le istruzioni. La responsabilit  si attribuisce senza difficolt : agli ingegneri per errori di specifica, ai committenti per scelte di deployment, agli utenti per usi impropri. Il sistema non   un agente in nessun senso filosoficamente rilevante:   uno strumento che amplifica e trasferisce l'agentivit  umana.

Il secondo tipo   quello dei sistemi di apprendimento automatico che ottimizzano funzioni obiettivo in modo emergente: i risultati del sistema non erano esplicitamente specificati dagli ingegneri, ma emergono dall'interazione tra la struttura dell'algoritmo, i dati di addestramento, e l'ambiente di deployment. Nessun singolo soggetto umano ha originariet  causale piena rispetto al comportamento emergente. Questo   il caso in cui il modello binario del soggetto nominativo produce il corto circuito che il quadro multidimensionale scioglie: invece di cercare il soggetto nominativo unico, si mappa la

distribuzione delle quattro dimensioni attraverso la catena causale, e si attribuiscono risposte reattive differenziate a ciascun attore in proporzione alle dimensioni che lo riguardano.

11.7 L'intelligenza artificiale come agente senza soggetto nominativo

Il problema più recente e forse più radicalmente nuovo che la struttura concettuale giuridica deve affrontare riguarda la responsabilità per danni causati da sistemi di intelligenza artificiale autonomi. Questo problema ha già generato una letteratura giuridica considerevole, ma essa ruota in gran parte attorno a una domanda mal posta: «chi è responsabile dei danni causati dall'AI?». La domanda è mal posta non perché non abbia risposta, ma perché presuppone che vi sia un soggetto nominativo — un «chi» — cui attribuire la responsabilità nel senso ordinario.

I sistemi di AI avanzati pongono il problema nella sua forma più pura: essi hanno alta originarietà causale (l'AI ha causato il danno), bassa o nulla intenzionalità nel senso umano (l'AI non «vuole» nel senso in cui vuole un essere umano), variabile controllabilità (a seconda del sistema e del contesto), e una responsabilità attribuibile che non può risiedere nell'AI stessa — che non è un soggetto giuridico — ma deve essere distribuita tra sviluppatore, produttore, distributore e utilizzatore secondo criteri che variano con le circostanze.

La proposta frequente di attribuire personalità giuridica all'AI — sulla base del modello della persona giuridica — è un tentativo di risolvere il problema del soggetto nominativo mancante attraverso la finzione giuridica. Ma essa non risolve il problema concettuale sottostante: anche se l'AI avesse personalità giuridica, rimane la questione di quale dimensione dell'agentività si fonda la sua responsabilità, e questa questione richiede la distinzione tra le dimensioni che la struttura binaria non offre.

Dal punto di vista dimensionale, la responsabilità per danni da AI è distribuita tra più soggetti secondo un pattern che corrisponde alla distribuzione delle dimensioni. Lo sviluppatore dell'algoritmo è originatore causale delle capacità dell'AI in senso remoto; ha intenzionalità rispetto al design ma non rispetto agli usi specifici; ha controllo sulle scelte di progettazione ma non sull'impiego; la responsabilità attribuibile per i danni da design difettoso è sua in misura proporzionale al

grado in cui il danno era prevedibile dal punto di vista del design. Il produttore del sistema ha originarietà causale nell'implementazione concreta; ha controllo sulla qualità del sistema così come distribuito; la sua responsabilità attribuibile è legata al controllo esercitato in quella fase. L'utilizzatore ha originarietà causale nel senso che ha scelto di impiegare il sistema in un certo contesto; ha intenzionalità rispetto all'impiego; ha controllo sull'uso specifico; la sua responsabilità è proporzionale alla prevedibilità del danno nell'uso specifico che ha fatto del sistema. Ogni posizione nella catena corrisponde a una configurazione dimensionale diversa, e ciascuna giustifica una forma di responsabilità qualitativamente diversa, non semplicemente un grado diverso di una stessa responsabilità.

II. La teoria del diritto e il problema della norma

11.8 La responsabilità come norma e come fatto: Kelsen e Hart rivisitati

Le grandi tradizioni della teoria del diritto del Novecento — il normativismo kelseniano e il positivismo analitico di Herbert Hart — hanno sviluppato concezioni della responsabilità giuridica che riflettono, ciascuna in modo diverso, le presupposizioni grammaticali identificate in questo testo.

Kelsen in *La dottrina pura del diritto* tratta la responsabilità come una categoria normativa: un soggetto è responsabile quando una norma giuridica attribuisce a lui, come condizione per l'applicazione di una sanzione, il fatto illecito. La responsabilità è definita in termini di imputazione normativa — *Zurechnung* — non in termini di agentività fattuale. Questa prospettiva ha il vantaggio di chiarire la struttura logica del giudizio di responsabilità, separandola dalle questioni empiriche sulla causalità e sull'intenzionalità. Ma essa ha il limite di non offrire criteri per valutare quali norme di imputazione siano giustificate: se la responsabilità è puramente normativa, qualsiasi norma di imputazione sembra ugualmente valida, e non si può rispondere alla domanda perché certi sistemi di responsabilità siano più giusti di altri.

Hart in *Causation in the Law* (con Tony Honore) e in *Punishment and Responsibility* ha sviluppato una teoria della responsabilità che cerca di connettere la struttura normativa con le strutture

fattuali dell'agentività. La nozione hartiana di «responsabilità come imputabilità» (*liability as ascriptivity*) identifica la responsabilità con l'appropriatezza di certe pratiche sociali rispetto a certi tipi di fatti riguardanti l'agente — in modo analogo, anche se non identico, alla proposta strawsoniana. Hart distingue tra tipi diversi di responsabilità — responsabilità causale, responsabilità per capacità, responsabilità come ruolo, responsabilità come imputazione giuridica — e mostra che ciascuno richiede condizioni diverse.

La teoria hartiana è più ricca della kelseniana rispetto alla complessità effettiva dell'agentività, ma rimane strutturalmente limitata dalla stessa binarietà che si è identificata in Strawson: Hart distingue tra presenza e assenza delle condizioni di responsabilità, senza articolare sistematicamente la multidimensionalità di quelle condizioni. La distinzione tra responsabilità causale e responsabilità per ruolo è una distinzione tra configurazioni dimensionali diverse — la prima si fonda prevalentemente sull'originarietà causale; la seconda sulla responsabilità attribuibile in senso istituzionale — ma Hart non la formula in questi termini, e la sua teoria non fornisce criteri per risolvere i casi in cui le due si separano.

Implicazioni per la Filosofia Politica

I. La responsabilità collettiva e le sue strutture

12.1 Il problema della responsabilità collettiva come problema dell'agentività distribuita

La filosofia politica affronta il problema dell'agentività in una forma che la filosofia morale individuale non ha — e che il quadro grammaticale binario è strutturalmente inadatto a trattare: la forma dell'**agentività collettiva e distribuita**. I fenomeni politici che richiedono attribuzione di responsabilità — la colonizzazione, il genocidio, la costruzione di strutture di oppressione sistemica, il cambiamento climatico come danno collettivo — non hanno agenti individuali identificabili nel senso in cui li hanno gli illeciti penali ordinari. Essi sono prodotti da milioni di azioni individuali, da strutture istituzionali, da abitudini culturali, da sistemi di incentivi:

l'attribuzione di responsabilità per questi fenomeni deve essere, per definizione, una responsabilità distribuita attraverso dimensioni diverse tra agenti diversi.

La filosofia politica anglosassone ha cercato di affrontare questo problema attraverso la nozione di **responsabilità collettiva** (*collective responsibility*), con esiti concettuali che riflettono le stesse difficoltà identificate nella filosofia morale individuale. La domanda «i tedeschi contemporanei sono responsabili dei crimini del nazismo?» — nella forma in cui Karl Jaspers la pose nel 1946 in *La questione della colpa* — o la domanda «i bianchi americani di oggi sono responsabili dell'eredità della schiavitù?» non trovano risposta soddisfacente nella struttura binaria. O vi è responsabilità — e allora si deve imputare ai discendenti ciò che non hanno compiuto — oppure non vi è — e allora si nega qualsiasi connessione tra una comunità e la propria storia. Entrambe le risposte sono insoddisfacenti, e lo sono perché la domanda è mal posta nella sua forma binaria.

Il quadro multidimensionale permette di riformulare la domanda in modo produttivo. La questione non è «i tedeschi contemporanei sono responsabili?» ma «rispetto a quali dimensioni dell'agentività vi è una connessione tra i tedeschi contemporanei e i crimini del nazismo?» L'originarietà causale assoluta è assente (i contemporanei non hanno compiuto quei crimini). L'intenzionalità è assente per definizione (non si può avere intenzione rispetto ad azioni non proprie). Ma vi può essere originarietà causale relativa in senso strutturale — le istituzioni, le strutture giuridiche, i vantaggi economici ereditati sono stati prodotti da quegli eventi e persistono nel presente. Vi può essere controllabilità attuale — il potere di trasformare o mantenere quelle strutture. Vi è certamente responsabilità attribuibile in senso politico — l'essere il referente appropriato di certe pratiche di riconoscimento, riparazione, cambiamento istituzionale.

12.2 Iris Marion Young e la responsabilità per le strutture: un'affinità teorica

Iris Marion Young in *Responsibility for Justice* (2011) — pubblicato postumo — ha sviluppato una teoria della responsabilità politica che ha forti affinità con l'approccio multidimensionale proposto in questo testo, sia pure senza il supporto dell'analisi grammaticale che qui si avanza. Young

distingue tra ciò che chiama il **liability model** e il **social connection model** della responsabilità.

Il liability model — la responsabilità come imputazione di un danno a un agente identificabile che ha agito in modo illecito — è il modello adeguato per i reati ordinari: funziona quando vi è un agente con piena intenzionalità, pieno controllo, originarietà causale diretta, che ha violato una norma esplicita. Ma esso fallisce sistematicamente quando si tenta di applicarlo ai danni strutturali prodotti da ingiustizie sistemiche: povertà, discriminazione, esclusione sociale. In questi casi non vi è un agente intenzionale del danno identificabile; il danno è il prodotto di milioni di interazioni che seguono le regole; nessuna singola azione è illecita nel senso che il liability model richiede.

Il social connection model di Young propone che la responsabilità per le strutture si fondi sulla partecipazione alle strutture stesse: chiunque contribuisca a riprodurre strutture ingiuste — anche senza intenzione e senza violazione di norme esplicite — ha una forma di responsabilità per quelle strutture che non è la responsabilità del liability model ma non è neppure assenza di responsabilità. Questa responsabilità è forward-looking piuttosto che backward-looking: non riguarda il passato in cui si è contribuito al danno ma il futuro in cui si può contribuire al cambiamento.

La proposta di Young è brillante e filosoficamente motivata, ma soffre di un'indeterminatezza che il quadro dimensionale può aiutare a risolvere. Il social connection model non specifica quale dimensione dell'agentività fonda la responsabilità strutturale: Young oscilla tra la partecipazione causale (si è parte delle strutture che producono il danno) e il controllo potenziale (si ha il potere di contribuire al cambiamento). Questi sono, nella terminologia di questo testo, due dimensioni diverse — l'originarietà causale relativa e la controllabilità — che producono forme diverse di responsabilità attribuibile.

Dal punto di vista della quadripartizione dimensionale, la responsabilità strutturale di Young può essere articolata nel modo seguente. Vi è una forma di responsabilità che si fonda sull'originarietà causale strutturale: contribuire a riprodurre strutture ingiuste, anche senza intenzione, fa di un agente una delle cause distribuite di quell'ingiustizia. Questa forma di responsabilità è analoga alla responsabilità oggettiva del diritto: si fonda sulla causalità, non sull'intenzionalità. Vi è poi una forma di responsabilità che si fonda sul controllo potenziale:

chi ha il potere di trasformare strutture ingiuste e non lo esercita ha una responsabilità per omissione che si fonda sul controllo. Questa forma è diversa dalla precedente: richiede non solo la partecipazione causale ma anche la disponibilità del controllo. E vi è infine una responsabilità attribuibile in senso politico che si fonda sull'essere membro della comunità politica che ha prodotto e mantiene quelle strutture: questa è la responsabilità che Jaspers chiamava «colpa politica» — distinta dalla colpa morale individuale ma non per questo meno reale.

12.3 Il problema della responsabilità storica: nazioni e eredità

Il problema della responsabilità delle nazioni per la propria storia — il tema della «transitional justice», delle commissioni di verità e riconciliazione, delle riparazioni per crimini storici — è il luogo in cui la filosofia politica si confronta più direttamente con la struttura multidimensionale dell'agentività collettiva.

Le commissioni di verità e riconciliazione — nella forma sviluppata in Sudafrica, Argentina, Ruanda, Sierra Leone e altrove — operano implicitamente con una concezione multidimensionale della responsabilità anche se non la formalizzano esplicitamente. Esse distinguono tra la responsabilità dei perpetratori diretti (alta in tutte le dimensioni), la responsabilità dei beneficiari delle strutture di oppressione (fondata sull'originarietà causale strutturale, non sull'intenzionalità), la responsabilità delle istituzioni che hanno prodotto o tollerato l'oppressione (fondata sul controllo istituzionale non esercitato), e la responsabilità della comunità nel suo insieme (una forma di responsabilità attribuibile collettiva che si fonda sull'appartenenza piuttosto che sull'agentività individuale).

Queste distinzioni pratiche, sviluppate nelle istituzioni di giustizia transizionale, sono più sofisticate della tassonomia strawsoniana: esse riconoscono implicitamente che le dimensioni dell'agentività si distribuiscono diversamente tra diversi tipi di agenti rispetto agli stessi eventi, e che questo richiede risposte di responsabilità qualitativamente diverse. Ma questa sofisticazione pratica non è stata tradotta in una teoria sistematica che le fondiamenti e ne articoli le implicazioni.

La questione delle **riparazioni** — il dibattito americano sulle riparazioni per la schiavitù è il caso più discusso ma non il solo — è il luogo in cui le conseguenze pratiche della struttura

dimensionale si manifestano con maggiore concretezza. I contrari alle riparazioni usano tipicamente l'argomento binario: i discendenti degli schiavisti non hanno commesso la schiavitù, dunque non sono responsabili di essa, dunque non devono riparazioni. I favorevoli usano argomenti che si riferiscono implicitamente a dimensioni diverse: i benefici economici ereditati (originarietà causale strutturale), la persistenza delle strutture di discriminazione che la schiavitù ha prodotto (controllabilità attuale delle strutture), l'identità comunitaria che include quella storia (responsabilità attribuibile politica).

Entrambi i lati del dibattito starebbero argomentando meglio se disponessero del quadro multidimensionale: i contrari dovrebbero specificare rispetto a quale dimensione negano la responsabilità (l'intenzionalità, certo; ma non necessariamente il vantaggio strutturale o il controllo delle strutture persistenti); i favorevoli dovrebbero distinguere quale forma di responsabilità fondano su quale dimensione, e quali pratiche di responsabilità sono appropriate a ciascuna dimensione.

12.4 Responsabilità politica e leadership: il problema dell'agentività mediata

La responsabilità dei leader politici per le conseguenze delle proprie politiche è un caso particolarmente istruttivo di agentività mediata. Un primo ministro che adotta una politica economica che produce, dieci anni dopo, un aumento della povertà, ha una configurazione dimensionale dell'agentività rispetto a quell'esito che è genuinamente complessa. L'originarietà causale è presente ma remota e mediata da catene causali lunghe e non interamente prevedibili. L'intenzionalità rispetto alla politica era presente; l'intenzionalità rispetto all'aumento di povertà era tipicamente assente o avversata. Il controllo sulla politica era alto; il controllo sulle conseguenze a lungo termine era basso per ragioni strutturali. La responsabilità attribuibile è presente in senso politico — il leader risponde politicamente delle proprie scelte — ma la sua natura è diversa dalla responsabilità morale individuale.

Max Weber ha articolato, in «La politica come professione» (1919), una distinzione che è direttamente rilevante per questa analisi: la distinzione tra **etica della convinzione** (*Gesinnungsethik*) e **etica della responsabilità** (*Verantwortungsethik*). Chi segue l'etica della convinzione è responsabile soltanto della purezza delle proprie intenzioni: risponde delle proprie azioni rispetto a un principio assoluto,

indipendentemente dalle conseguenze. Chi segue l'etica della responsabilità risponde anche delle conseguenze prevedibili delle proprie azioni, anche quelle non intenzionate.

Questa distinzione weberiana è, nella terminologia di questo testo, una distinzione tra due concezioni di quale dimensione dell'agentività fonda la responsabilità politica: l'intenzionalità (per l'etica della convinzione) oppure l'originarietà causale relativa combinata con il controllo (per l'etica della responsabilità). Weber sostiene che la politica richieda l'etica della responsabilità, non quella della convinzione, perché chi detiene il potere non può limitare la propria responsabilità all'intenzione: è anche responsabile per le conseguenze di azioni compiute con buona intenzione che producono effetti dannosi prevedibili.

La distinzione weberiana è filosoficamente perspicace, ma la sua formulazione rimane binaria: o si segue l'etica della convinzione oppure quella della responsabilità. Il quadro multidimensionale permette di articolare questa distinzione in modo più preciso: non si tratta di scegliere tra intenzione e conseguenza come fondamento unico della responsabilità, ma di specificare per ogni dimensione quale tipo di responsabilità fonda e quali pratiche di attribuzione giustifica. Un leader politico risponde delle proprie intenzioni in senso morale; risponde delle conseguenze prevedibili delle proprie azioni in senso politico-istituzionale; risponde delle strutture che mantiene o non trasforma in senso di responsabilità strutturale. Queste tre forme di responsabilità richiedono pratiche di attribuzione diverse e non sono riducibili l'una all'altra.

II. Libertà politica e strutture di coazione

12.5 La libertà negativa e il soggetto nominativo

Isaiah Berlin in «Due concetti di libertà» (1958) ha articolato la distinzione tra libertà negativa — assenza di impedimenti esterni all'azione — e libertà positiva — il possesso delle capacità e delle risorse necessarie per realizzare i propri scopi. Questa distinzione è diventata uno dei punti di riferimento canonici della filosofia politica liberale, e ha generato una letteratura vastissima.

La distinzione berliniana è costruita attorno a una concezione dell'agente come soggetto nominativo: il soggetto che agisce o è impedito dall'agire. La libertà negativa misura l'assenza di impedimenti rispetto all'azione del soggetto; la libertà positiva misura le capacità del soggetto di perseguire i propri scopi. In entrambi i casi, il soggetto è l'unità di analisi — l'agente con le sue azioni, i suoi impedimenti, le sue capacità.

Ma la distinzione berliniana non articola le dimensioni dell'agentività che questo testo identifica, e questa mancanza produce lacune teoriche che la letteratura successiva ha faticato a colmare. In particolare, Berlin non distingue tra i diversi tipi di impedimento alla libertà che corrispondono alle diverse dimensioni dell'agentività. Un impedimento alla libertà può operare sulla dimensione dell'intenzionalità — manipolazione delle preferenze, condizionamento ideologico — senza produrre impedimenti fisici all'azione. Può operare sulla dimensione del controllo — strutture sociali che rendono certi comportamenti *de facto* impossibili senza proibirli formalmente. Può operare sulla dimensione della responsabilità attribuibile — sistemi che rendono l'agente responsabile di eventi che non origina causalmente né controlla. Queste forme di riduzione della libertà sono concettualmente distinte, ma la struttura berliniana non le distingue: esse rientrano tutte, imprecisamente, nella categoria generale della libertà positiva o rimangono fuori dalla struttura concettuale.

Philip Pettit in *Republicanism* (1997) ha proposto un terzo concetto di libertà — la libertà come non-dominazione — che cerca di catturare forme di riduzione della libertà che la distinzione berliniana lascia nell'ombra. La dominazione, per Pettit, non è semplicemente l'interferenza effettiva nell'azione: è la condizione di essere soggetti al potere arbitrario di un altro, il quale potrebbe interferire anche se non lo fa attualmente. Questa proposta è più fine di quella berliniana, ma rimane anch'essa strutturalmente limitata: la non-dominazione è una proprietà binaria — un agente è o non è in condizione di dominazione — senza articolare le dimensioni dell'agentività rispetto alle quali la dominazione opera.

12.6 La teoria della giustizia e le capacità

L'approccio delle capacità — sviluppato da Amartya Sen in *Inequality Reexamined* (1992) e Martha Nussbaum in *Women and Human Development* (2000) — rappresenta il tentativo più sofisticato nella filosofia politica contemporanea di articolare

una concezione multidimensionale del benessere e della libertà che superi le limiti dell'approccio utilitarista e rawlsiano.

Sen propone di valutare la libertà e il benessere non in termini di utilità o di risorse, ma in termini di **capacità** (*capabilities*): le opportunità reali di vita che un individuo ha, la gamma delle vite che può effettivamente condurre. Questa prospettiva riconosce implicitamente che la libertà di cui le teorie politiche devono occuparsi non è una sola cosa: è un insieme di capacità distinte che possono variare indipendentemente.

L'affinità tra l'approccio delle capacità e l'approccio multidimensionale proposto in questo testo è strutturale. L'approccio delle capacità dice, in sostanza, che il soggetto politico non è un'entità omogenea con un'unica proprietà (utilità, risorse, libertà in senso binario) ma un'entità multidimensionale le cui libertà rilevanti devono essere articolate separatamente. Questo corrisponde esattamente alla tesi che l'agentività del soggetto morale non è un'unica proprietà binaria ma un insieme di dimensioni distinte.

Nussbaum identifica dieci capacità centrali — vita, salute corporea, integrità corporea, sensi/immaginazione/pensiero, emozioni, ragione pratica, affiliazione, altre specie, gioco, controllo del proprio ambiente — e sostiene che ciascuna ha importanza intrinseca e non è sostituibile da un ammontare maggiore delle altre. Questa incommensurabilità delle capacità è analoga all'indipendenza delle dimensioni dell'agentività: non si può compensare la mancanza di controllo con più originarietà causale, così come non si può compensare la mancanza di integrità corporea con più affiliazione.

Ma l'approccio di Nussbaum, che lavora sul lato del benessere e delle libertà positive, non si connette sistematicamente con la teoria della responsabilità e degli atteggiamenti reattivi che è al centro del presente testo. La connessione è questa: se le capacità centrali sono incommensurabili e multidimensionali, allora le violazioni di quelle capacità — le azioni e le strutture che le compromettono — devono essere valutate rispetto alle dimensioni dell'agentività in modo multidimensionale. Un'azione che compromette l'integrità corporea con piena intenzionalità e pieno controllo è diversa da un'azione che la compromette come effetto collaterale non intenzionale di una struttura che riduce le capacità di accesso ai servizi sanitari. Entrambe richiedono risposte politiche, ma risposte qualitativamente diverse che corrispondono a configurazioni dimensionali diverse dell'agentività di chi ha prodotto il danno.

III. La politica della responsabilità: populismo, scapegoating, e distorsioni

12.7 La semplificazione binaria come strumento politico

La struttura binaria dell'agentività — qualcuno è responsabile o non è responsabile, qualcuno è colpevole o è innocente — non è solo un limite filosofico: è anche uno strumento politico. La binarietà dell'attribuzione della responsabilità è funzionale alla semplificazione politica, che a sua volta è funzionale alla mobilitazione delle emozioni. Il populismo, nelle sue forme diverse, sfrutta sistematicamente la struttura binaria: identifica un agente responsabile di tutti i mali (gli immigrati, le élite, le banche, i burocrati di Bruxelles), attribuisce a quell'agente piena originarietà causale, piena intenzionalità, pieno controllo, e piena responsabilità attribuibile rispetto a fenomeni che hanno invece strutture causali complesse e distribuite.

Questa semplificazione non è accidentale: essa è resa possibile dalla struttura grammaticale binaria dell'agentività, che fornisce le categorie cognitive automatiche su cui la semplificazione politica fa leva. Il populismo non inventa la binarietà: la eredita dalla grammatica e la amplifica. Quando si dice «i banchieri hanno rovinato l'economia» o «i burocrati europei controllano le nostre vite», si usa la struttura grammaticale del soggetto nominativo — un agente identificabile che compie un'azione identificabile — per descrivere fenomeni che sono invece il risultato di strutture sistemiche, di interazioni tra milioni di scelte, di incentivi istituzionali, di vincoli strutturali.

L'analisi dimensionale offre uno strumento critico per smontare queste semplificazioni senza ricorrere all'argomento opposto e ugualmente semplificato che «nessuno è responsabile». La risposta non è «nessuno è responsabile della crisi finanziaria» ma «la responsabilità si distribuisce in modo complesso rispetto a dimensioni diverse tra agenti diversi: i banchieri che hanno creato prodotti finanziari opachi hanno alta originarietà causale e alta intenzionalità rispetto a quelle scelte; i regolatori che non hanno esercitato il controllo disponibile hanno responsabilità fondata sulla controllabilità non esercitata; i politici che hanno deregolamentato il settore hanno responsabilità strutturale fondata sulle strutture di incentivi create». Questa risposta è più complessa di quella

populista, ma è più vera, e offre anche una base più adeguata per le politiche di risposta e di riforma.

12.8 Il problema del «capro espiatorio» e la giustizia retributiva

René Girard ha elaborato una teoria del capro espiatorio come meccanismo fondamentale di coesione sociale: le comunità riducono la violenza interna dirigendola verso un agente designato come colpevole, indipendentemente dalla sua effettiva responsabilità. Questa teoria, sviluppata principalmente in chiave antropologica e religiosa, ha implicazioni filosofico-politiche dirette che l'analisi dimensionale permette di precisare.

Il meccanismo del capro espiatorio è, nella terminologia di questo testo, un meccanismo di distorsione della responsabilità attribuibile: esso attribuisce a un agente la responsabilità attribuibile rispetto a danni di cui quell'agente non aveva originarietà causale, né intenzionalità, né controllo. La funzione sociale del meccanismo — produrre coesione attraverso la concentrazione degli atteggiamenti reattivi negativi su un bersaglio designato — è reale, ma essa si realizza precisamente attraverso l'ignoranza delle altre dimensioni dell'agentività: il capro espiatorio funziona perché la struttura grammaticale binaria rende naturale l'attribuzione di piena responsabilità a un agente identificabile, senza che la distinzione tra le dimensioni impedisca la semplificazione.

La risposta istituzionale al meccanismo del capro espiatorio — il sistema del due process nel diritto, i meccanismi di verifica delle responsabilità nelle istituzioni politiche, le pratiche di revisione storica — può essere compresa come il tentativo istituzionale di resistere alla semplificazione binaria e di imporre la distinzione tra le dimensioni. Un sistema giuridico in cui la responsabilità penale richiede la prova dell'intenzionalità (*mens rea*) è un sistema che rifiuta di fondare la responsabilità sulla sola originarietà causale: esso insiste che non basta aver causato il danno, occorre averlo causato con il grado appropriato di intenzionalità. Questo non è arbitrario: è la traduzione istituzionale dell'intuizione che le dimensioni dell'agentività sono distinte e che la responsabilità attribuibile varia a seconda di quale configurazione dimensionale è presente.

Case di Studio Storico-Filosofici

I. La grammatica come storia del pensiero

13.1 Il metodo dei case studies storico-filosofici

I casi di studio presentati in questa sezione hanno un doppio scopo. Il primo è verificativo: essi cercano evidenza storica del fatto che la struttura grammaticale delle lingue in cui certe tradizioni filosofiche si sono formate abbia lasciato tracce riconoscibili nel modo in cui queste tradizioni hanno posto il problema dell'agentività e della responsabilità. Il secondo è costruttivo: essi mostrano come un diverso sistema grammaticale abbia talvolta prodotto soluzioni al problema dell'agentività che anticipano, o convergono con, la struttura multidimensionale proposta in questo testo — soluzioni che però non sono state incorporate nella tradizione dominante perché quella tradizione si è formata in un contesto grammaticale diverso.

Il metodo è, come si è dichiarato nella prima parte, storico-ricostruttivo: esso non dimostra relazioni causali tra grammatica e filosofia, ma identifica corrispondenze strutturali che sono interpretabili come tracce dell'influenza grammaticale. Le conclusioni di questa sezione sono interpretative, non dimostrative, e si reggono sulla qualità dell'analisi testuale più che su prove di tipo sperimentale.

II. La filosofia greca classica

13.2 Aristotele e le dimensioni dell'agentività nell'Etica Nicomachea

L'Etica Nicomachea di Aristotele contiene, in forma non sistematizzata ma filosoficamente precisa, le distinzioni dimensionali dell'agentività che la tradizione successiva ha perso nella transizione verso le lingue romanze e germaniche. Il fatto che queste distinzioni esistano in Aristotele ma non siano state ereditate dalla tradizione che si richiama ad Aristotele è

uno dei più istruttivi esempi dell'effetto di filtro grammaticale che questo testo ha cercato di documentare.

Il libro III dell'Etica Nicomachea analizza le condizioni dell'azione volontaria con una ricchezza che ha pochi equivalenti nella tradizione successiva. Aristotele distingue tra azioni **hekon** (volontarie), **akon** (involontarie) e **ouch hekon** (non-volontarie, una categoria distinta dall'involontario). Questa tripartizione — che traduce quasi automaticamente in italiano come «volontario/involontario» perdendo la terza categoria — è già più ricca della struttura binaria che domina la tradizione analitica.

L'involontario (*akon*) comprende per Aristotele due tipi di azione radicalmente diversi: le azioni compiute **per costrizione** (*biai*) — in cui un agente esterno causa fisicamente l'azione dell'agente — e le azioni compiute **per ignoranza** (*di' agnoian*) — in cui l'agente non conosce i fatti rilevanti dell'azione che compie. Questa distinzione è filosoficamente precisa: la costrizione fisica opera sulla dimensione del controllo (l'agente non controllava l'azione); l'ignoranza opera sulla dimensione dell'intenzionalità (l'agente non aveva le informazioni che avrebbero reso intenzionale rispetto all'esito rilevante l'azione compiuta). Aristotele separa questi due casi perché le conseguenze per la responsabilità e per il rimpianto (*metameleia*) sono diverse: chi ha agito per costrizione non si rammarica dell'azione nel modo in cui si rammarica chi ha agito per ignoranza. Questa differenziazione della risposta emotiva in funzione della dimensione dell'agentività compromessa è esattamente la struttura della tassonomia differenziata degli atteggiamenti reattivi proposta in questo testo.

Ancora più rilevante è la distinzione aristotelica tra l'ignoranza **della** propria azione — non sapere cosa si sta facendo — e l'ignoranza **nelle circostanze** dell'azione — non sapere certi fatti rilevanti del contesto. Solo la seconda produce involontarietà vera; la prima — la mancanza di consapevolezza dovuta a ubriachezza o a vizio morale — non scusa perché l'agente era originatore causale dello stato di ignoranza in cui si trovava. Questa distinzione aristotelica grammaticalizza avant la lettre la differenza tra l'originarietà causale dell'ignoranza stessa — chi ha prodotto lo stato di non-sapere — e l'ignoranza come condizione che opera sull'intenzionalità.

La categoria della **non-volontarietà** (*ouch hekon*) — che Aristotele distingue dall'involontarietà (*akon*) — comprende azioni come quelle compiute in circostanze in cui la scelta è

formalmente disponibile ma il contesto la rende di fatto obbligata: gettare le merci in mare durante una tempesta per salvare la nave e l'equipaggio è, per Aristotele, un'azione che si compie *hekon* nel momento in cui la si compie — perché si sceglie deliberatamente — ma che ha carattere *akon* in senso più ampio — perché nessuno avrebbe fatto quella scelta se non si fosse in quelle circostanze. Questa categoria è la descrizione di una configurazione dimensionale specifica: alta intenzionalità nell'atto singolo, alta controllabilità dell'esecuzione, ma ridotta libertà della situazione complessiva in cui la scelta avviene. Il diritto moderno ha una categoria analoga nella costrizione situazionale (*stato di necessità*), ma la sua giustificazione è meno articolata concettualmente di quanto non lo sia la tripartizione aristotelica.

13.3 La voce media greca: un dispositivo grammaticale per l'agentività intermedia

Un elemento grammaticale del greco antico che ha implicazioni dirette per la struttura dell'agentività e che è andato perduto nella tradizione filosofica occidentale con la transizione verso le lingue romanze è la **voce media**. Il greco classico ha tre voci verbali: l'attivo, il medio e il passivo. Mentre l'attivo e il passivo corrispondono abbastanza direttamente alle strutture attiva e passiva delle lingue romanze, la voce media non ha equivalente morfologico nelle lingue europee moderne: essa è scomparsa come categoria grammaticale autonoma, e le sue funzioni sono distribuite tra diverse strutture lessicali o sintattiche.

La voce media greca esprime un'ampia gamma di situazioni in cui l'agente è al tempo stesso soggetto e paziente dell'azione: azioni in cui il soggetto agisce su se stesso, azioni in cui il soggetto agisce nel proprio interesse, azioni in cui il soggetto è la sede di un processo che origina da lui ma che egli non governa pienamente, stati in cui il soggetto è caratterizzato da un processo piuttosto che da una azione. Questa gamma semantica è molto più ricca di ciò che in italiano si esprime con i verbi riflessivi: il medio greco grammaticalizza una zona di agentività intermedia tra il pieno controllo attivo e la piena passività ricevente.

Per la filosofia dell'azione, la voce media è particolarmente significativa perché essa grammaticalizzava proprio la classe di eventi che producono le maggiori difficoltà nella struttura binaria: gli eventi in cui l'agente origina l'azione ma non la controlla pienamente, in cui l'azione «accade a» l'agente nel

momento stesso in cui l'agente la «produce», in cui l'intenzionalità e l'originarietà causale sono presenti ma il controllo è ridotto o assente. La filosofia greca, avendo a disposizione questa risorsa grammaticale, poteva esprimere queste sfumature con più naturalezza di quanto non possa fare la tradizione filosofica delle lingue romanze e germaniche.

La perdita della voce media con la fine del greco classico come lingua viva del pensiero filosofico — sostituita dal latino, che ne aveva solo tracce residuali, e poi dalle lingue romanze e germaniche, che ne avevano ancor meno — è un caso concreto dell'effetto di filtro grammaticale: una struttura linguistica che grammaticalizzava distinzioni filosoficamente rilevanti è scomparsa dal patrimonio grammaticale della tradizione dominante, e le distinzioni che essa rendeva naturali sono diventate più difficili da formulare.

La tradizione analitica contemporanea ha cercato di recuperare queste distinzioni attraverso l'analisi filosofica deliberata — le discussioni sulla voce media nella filosofia dell'azione, la distinzione tra azioni in cui il corpo «fa» e azioni che l'agente «fa», l'analisi delle azioni automatiche e delle abitudini — ma lo ha fatto con molto più sforzo di quanto sarebbe necessario se la struttura grammaticale fornisse automaticamente le categorie.

III. La filosofia scolastica e la libertà della volontà

13.4bis Duns Scoto e la contingenza radicale della volontà: un'anticipazione medievale

Il dibattito medievale sulla libertà della volontà non si esaurisce nella triade agostiniana che il testo esamina. Duns Scoto ha sviluppato, in risposta critica sia ad Agostino sia a Tommaso, una teoria della libertà della volontà che anticipa in modo sorprendente alcune delle questioni che il quadro multidimensionale mette a fuoco. Per Scoto, la libertà della volontà non consiste nell'essere determinata dalla ragione — come nella concezione tomistica in cui l'intelletto illumina la volontà e la volontà segue il bene conosciuto — né nell'essere liberata dalla grazia — come nella concezione agostiniana. Essa consiste in una capacità intrinseca alla volontà di determinare se stessa, indipendentemente sia dalla necessità naturale sia dalla sufficienza della ragione: la volontà può scegliere il bene minore anche quando il bene maggiore è perfettamente conosciuto,

non per debolezza o ignoranza, ma per la pura contingenza della propria auto-determinazione.

Questa tesi scotista è la formulazione medievale più precisa di ciò che Galen Strawson chiamerà nel ventesimo secolo il requisito della «self-origination»: per essere pienamente libero, l'agente deve essere la sorgente ultima delle proprie determinazioni, non il risultato di forze causali esterne. Scoto credeva che questa self-origination fosse possibile — era, per lui, la struttura della volontà umana come immagine della libertà divina —; Galen Strawson sostiene che sia impossibile. Il quadro multidimensionale permette di vedere che entrambi si concentrano su una sola dimensione dell'agentività — l'originarietà causale assoluta — e la trattano come condizione necessaria della responsabilità piena, ignorando le altre dimensioni che la responsabilità ordinaria richiede e che non presuppongono quella forma estrema di auto-determinazione.

Dal punto di vista grammaticale, Scoto scriveva in latino — una lingua con la stessa struttura nominativo-accusativa che questa trattazione esamina. La sua teoria della contingenza radicale della volontà è il tentativo di formulare in latino una struttura dell'agentività che il latino non grammaticalizza automaticamente: l'auto-determinazione del soggetto che sfugge sia alla necessità naturale sia alla determinazione razionale. La complessità e l'oscurità della teoria scotista — che ha generato secoli di commentari e controversie — è in parte la difficoltà di esprimere in una grammatica binaria una struttura che richiederebbe, per essere espressa con naturalezza, categorie grammaticali che il latino non possiede.

13.4 Il dibattito medievale sulla volontà: Agostino, Anselmo, Tommaso

Il dibattito medievale sulla libertà della volontà — *liberum arbitrium* — ha avuto luogo principalmente in latino, e la struttura grammaticale del latino ha co-determinato la forma dei problemi in modo interessante. Il latino ha perso le distinzioni della voce media greca, ma ha mantenuto una distinzione morfosintattica tra il verbo al modo indicativo e il verbo al modo congiuntivo che grammaticalizza, sia pure in modo obliquo, la distinzione tra ciò che accade effettivamente e ciò che potrebbe accadere. Questa distinzione modale è rilevante per il problema del libero arbitrio perché il problema controfattuale — «avrebbe potuto agire altrimenti?» — è un

problema di modo congiuntivo: richiede di ragionare su ciò che non è accaduto ma che avrebbe potuto accadere.

Agostino ha posto il problema della libertà della volontà in modo che riflette la struttura grammaticale latina con particolare chiarezza. Il suo problema centrale — come la volontà possa essere libera se Dio ha prescienza degli atti futuri — è formulato in termini che presuppongono la binarietà: o la volontà è libera (e allora Dio non sa in anticipo ciò che essa farà) oppure Dio conosce gli atti futuri (e allora la volontà non è libera). La soluzione agostiniana — che la prescienza divina è compatibile con la libertà umana perché Dio conosce la libera scelta futura senza determinarla — è una soluzione compatibilista ante litteram, ma formulata all'interno di una struttura binaria che non distingue le dimensioni dell'agentività.

Anselmo di Aosta, nel *De libertate arbitrii*, introduce una distinzione che ha un'affinità strutturale con la distinzione dimensionale proposta in questo testo. Anselmo distingue tra la libertà come **potere di non peccare** (*posse non peccare*) — la capacità di astenersi dal male — e la libertà come **impossibilità di peccare** (*non posse peccare*) — la condizione dell'agente che non può cedere al male, come gli angeli confermati in grazia. Questa distinzione non è una distinzione di grado lungo una scala unica: è una distinzione qualitativa tra due modi diversi di essere liberi che corrispondono a due configurazioni diverse delle dimensioni dell'agentività. Il potere di non peccare corrisponde alla libertà che si fonda sul controllo: l'agente ha la capacità di controllare la propria azione e di orientarla verso il bene. L'impossibilità di peccare corrisponde a qualcosa di più profondo: una trasformazione della struttura stessa dell'originarietà causale e dell'intenzionalità dell'agente, per cui il male non può più emergere come possibilità dalla struttura della sua volontà.

Tommaso d'Aquino ha sviluppato la distinzione anselmiana in una direzione che riflette l'influenza aristotelica e ha implicazioni per l'analisi dimensionale. Nella *Summa Theologiae*, Tommaso distingue tra gli **atti dell'uomo** (*actus hominis*) — gli atti che l'uomo compie come essere biologico, comuni agli animali, come respirare e digerire — e gli **atti umani** (*actus humani*) — gli atti che l'uomo compie in quanto specificamente umano, attraverso la ragione e la volontà. Solo gli atti umani sono soggetti alla valutazione morale.

Questa distinzione tommasiana ha una struttura analoga alla distinzione georgiana tra il soggetto dativo (locus dell'evento

biologico) e il soggetto ergativo (originatore causale intenzionale). Gli *actus hominis* corrispondono alla configurazione in cui il soggetto è il locus dell'evento fisiologico, non il suo originatore intenzionale: l'individuo che respira o digerisce non è l'agente di quelle azioni nel senso moralmente rilevante. Gli *actus humani* corrispondono alla configurazione in cui il soggetto è l'originatore intenzionale: l'individuo che delibera e sceglie è l'agente nel senso pieno.

Ma la distinzione tommasiana, pur cogliendo qualcosa di reale, rimane binaria: o l'atto è dell'uomo (non soggetto a valutazione morale) o è umano (soggetto a valutazione morale). Essa non articola le configurazioni intermedie che la morfologia georgiana grammaticalizza: gli atti che hanno originarietà causale fisiologica ma che l'agente può modulare con la ragione; gli atti che erano pienamente volontari in un momento anteriore ma sono diventati automatici attraverso l'abitudine; gli atti che originano da disposizioni acquisite volontariamente e ora non più sotto il controllo diretto. Tommaso discute queste configurazioni altrove — nella teoria dell'abito morale, nell'analisi della passione e della concupiscenza — ma senza integrarle nella distinzione fondamentale degli atti.

13.5 Il dibattito sulla grazia tra Erasmo e Lutero: una contesa grammaticale

La controversia tra Erasmo e Lutero sulla libertà del volere — *De libero arbitrio* di Erasmo (1524) contro il *De servo arbitrio* di Lutero (1525) — è uno degli episodi più drammatici della storia dell'idea di libertà, e ha avuto conseguenze politiche e culturali incalcolabili. Essa è anche, nella prospettiva di questo testo, un esempio istruttivo di come una disputa genuinamente filosofica si sia svolta all'interno di un quadro grammaticale che ne ha pre-determinato la struttura.

Erasmo difende la libertà del volere come compatibile con la grazia divina: la volontà umana, sia pure indebolita dal peccato originale, conserva una capacità di cooperare con la grazia, di orientarsi verso il bene con l'aiuto divino. Lutero nega radicalmente questa capacità: la volontà umana è servo del peccato, non ha libertà di autodeterminazione rispetto al bene soprannaturale. La grazia opera senza cooperazione della volontà.

La disputa si svolge interamente all'interno della struttura binaria: o la volontà è libera (e allora può cooperare) oppure è

schiaiva (e allora non può). Questa binarietà è riflessa grammaticalmente nel latino di entrambi i contendenti: la volontà è il soggetto nominativo che ha o non ha la proprietà della libertà. Ciò che manca nella disputa è precisamente la distinzione tra le dimensioni: la volontà potrebbe avere diversi gradi di libertà rispetto a dimensioni diverse — potrebbe conservare capacità di controllo in certi ambiti pur avendo perso originarietà causale rispetto al bene soprannaturale; potrebbe avere intenzionalità rispetto ai propri atti particolari pur non avendo libertà strutturale rispetto alla propria condizione complessiva.

Erasmus tenta implicitamente di introdurre distinzioni dimensionali — distinguendo tra il cooperare con la grazia e il meritare la salvezza, tra la volontà come capacità debole e la volontà come causa sufficiente — ma senza il quadro concettuale che gli permetterebbe di formalizzare queste distinzioni come distinzioni tra dimensioni dell'agentività. Lutero, da parte sua, ha un'intuizione profonda che corrisponde alla concessione incompatibilista sull'originarietà causale ultima: la volontà umana non è la sorgente causale ultima del proprio orientamento verso il bene, perché quella sorgente è Dio. Ma Lutero trasforma questa intuizione in una negazione totale della libertà della volontà che travolge anche le dimensioni del controllo, dell'intenzionalità e della responsabilità attribuibile — e questo è l'errore, precisamente dell'errore incompatibilista identificato in questo testo: dalla concessione sull'originarietà causale ultima non segue il collasso di tutte le altre dimensioni.

IV. La filosofia moderna e l'emergenza del soggetto

13.6 Cartesio e la riduzione dell'agentività al soggetto pensante

La rivoluzione cartesiana — la fondazione della certezza nella *res cogitans*, nel soggetto pensante come unico punto di accesso all'essere — ha implicazioni per la struttura dell'agentività che raramente vengono discusse in questo contesto. Cartesio radicalizza la struttura grammaticale del soggetto nominativo portandola all'estremo: il soggetto pensante è l'unica entità di cui si abbia certezza, e tutto il resto — il corpo, il mondo esterno, gli altri soggetti — è accessibile solo in modo mediato attraverso il pensiero.

Questa radicalizzazione ha un effetto diretto sulla struttura dell'agentività: essa colloca il soggetto pensante come sorgente di tutte le sue azioni nel senso più forte. Se il corpo è *res extensa* — pura meccanica — e se la mente è *res cogitans* — pura spontaneità pensante — allora le azioni del soggetto devono originarsi dalla mente, non dal corpo. Il problema della volontà e dell'azione diventa il problema di come la mente possa muovere il corpo, cioè di come la *res cogitans* possa causare eventi nella *res extensa*. Questo è il famoso problema dello *union* cartesiano, e la difficoltà che Cartesio ha nel risolverlo è direttamente rilevante per il tema dell'agentività.

Dal punto di vista della quadripartizione dimensionale, il cartesianesimo produce una concezione dell'agentività in cui l'originarietà causale è assolutamente localizzata nella mente, l'intenzionalità è un attributo esclusivo della mente, il corpo è mero esecutore meccanico senza agentività propria, e la responsabilità attribuibile si fonda interamente sulla mente. Questa struttura ha una coerenza interna, ma ha il costo di rendere incomprensibile la classe di eventi corporei che hanno un grado di agentività intermedio: gli atti automatici, le abitudini, le risposte emotive corporee, i movimenti che sono in parte intenzionali e in parte automatici. Tutte queste situazioni producono paradossi nel quadro cartesiano, perché esse non sono né puramente meccaniche (come la digestione) né puramente mentali (come la decisione deliberata).

Maurice Merleau-Ponty, nella *Fenomenologia della percezione* (1945), ha sviluppato la critica più profonda al dualismo cartesiano proprio attraverso l'analisi di queste situazioni intermedie: il corpo vissuto — il *corps propre* — non è né la meccanica cartesiana né il puro soggetto pensante, ma un'entità che ha una propria struttura intenzionale pre-riflessiva che grammaticalizza, per così dire, l'agentività corporea in modo irriducibile a entrambi i poli del dualismo. La fenomenologia del corpo vissuto di Merleau-Ponty è, nella terminologia di questo testo, una fenomenologia delle dimensioni dell'agentività corporea che il dualismo cartesiano non può articolare.

13.7 Kant e l'autonomia come originarietà causale assoluta

La filosofia morale kantiana è il luogo della tradizione occidentale in cui la struttura dell'originarietà causale assoluta — la dimensione che questo testo concede all'incompatibilismo — viene portata al centro dell'analisi morale in modo più

esplicito e sistematico. La nozione kantiana di **autonomia** — il darsi della legge a se stessi — è precisamente la nozione di originarietà causale assoluta applicata alla struttura della volontà morale.

Per Kant, l'agente morale è libero quando la legge che governa la sua azione origina da lui stesso come agente razionale, non da impulsi sensibili o da autorità esterne. L'autonomia è l'auto-nomia — il darsi la propria norma — e questo darsi-la-norma è la forma di originarietà causale più profonda che Kant possa immaginare: non l'originarietà causale all'interno della serie naturale, che è sempre relativa (ogni causa è effetto di una causa precedente), ma l'originarietà causale della ragione rispetto a se stessa — la spontaneità della ragion pratica pura.

Questa concezione kantiana è resa possibile dalla struttura del tedesco, che grammaticalizza la riflessività — il darsi-la-norma a se stessi — in modo più naturale del latino o dell'inglese, attraverso i verbi riflessivi che usano il proprio soggetto come oggetto. Ma la sua struttura filosofica profonda è quella del soggetto nominativo portato all'estremo: la ragione pura è il soggetto che non dipende da nulla di esterno a sé, che è originatore causale assoluto della propria legge.

Il problema del kantismo, identificato da Hegel e sviluppato da una lunga tradizione di critici, è che l'autonomia kantiana astrae l'agente dalla sua situazione storica, sociale e corporea: il soggetto morale kantiano non ha corpo, non ha storia, non ha relazioni di dipendenza con altri — è pura razionalità auto-sufficiente. Questa astrazione è il riflesso filosofico del soggetto grammaticale portato all'estremo: il soggetto nominativo privo di ogni determinazione materiale che non sia la propria spontaneità razionale.

La critica hegeliana all'etica formale kantiana — che la legge morale universale è vuota senza il contenuto che viene dalla *Sittlichkeit*, dall'eticità concreta delle istituzioni e dei costumi storicamente determinati — è, in parte, la critica alla struttura grammaticale dell'autonomia kantiana: essa mostra che l'originarietà causale assoluta del soggetto morale è un'astrazione, e che l'agentività concreta è sempre già strutturata da contesti relazionali, istituzionali e storici che non derivano dall'autonomia del soggetto.

13.8 Hegel e la dialettica dell'agentività: un tentativo parziale di superamento

Hegel costituisce il caso forse più interessante per l'analisi storico-grammaticale, perché la sua filosofia rappresenta il tentativo più sistematico, all'interno della tradizione idealista tedesca, di superare la struttura binaria dell'agentività senza disporre del quadro concettuale che il georgiano offre.

La Fenomenologia dello Spirito analizza l'agentività in modo che ha strutturale affinità con la quadripartizione dimensionale, sia pure in un idioma completamente diverso. La figura del servo e del signore grammaticalizza la distinzione tra originarietà causale e responsabilità attribuibile: il signore è il referente delle pratiche di riconoscimento — la responsabilità attribuibile è sua — ma l'originarietà causale delle trasformazioni materiali del mondo è del servo. Questa distinzione tra chi «ha» la responsabilità sociale e chi «produce» le trasformazioni causali è esattamente la distinzione tra la quarta e la prima dimensione dell'agentività.

La dialettica hegeliana della coscienza morale nel capitolo sull'«anima bella» e sulla «coscienza agente» articola la distinzione tra intenzionalità e originarietà causale in un modo che anticipa alcune delle difficoltà della tradizione analitica. L'anima bella — che si rifugia nell'intenzione pura senza agire, per non contaminare la propria purezza con le conseguenze incontrollabili dell'azione — è la figura di chi riduce la responsabilità morale alla sola dimensione dell'intenzionalità. L'hegelismo risponde che questa riduzione è insostenibile: l'agire produce sempre conseguenze che sfuggono all'intenzione, e la responsabilità dell'agente include anche queste conseguenze non intenzionate nella misura in cui il soggetto era l'originatore causale dell'azione.

La dialettica hegeliana ha però un limite strutturale rispetto alla quadripartizione dimensionale: essa tende a risolvere le tensioni tra le dimensioni in una sintesi superiore piuttosto che a mantenerle come distinzioni permanenti che producono forme diverse di responsabilità. Il sistema hegeliano ha una spinta teleologica verso l'integrazione — ogni opposizione deve essere superata in una sintesi — che non lascia spazio alla coesistenza permanente di dimensioni irriducibili. La struttura georgiana, al contrario, grammaticalizza le dimensioni come distinte in modo permanente: non vi è una sintesi che le unifica, vi è una costellazione di valori che variano indipendentemente.

V. La filosofia buddhista e la non-sé: un caso di grammatica alternativa

13.9 Il georgiano e il buddhismo: due soluzioni a un problema condiviso

Il buddhismo costituisce il case study filosofico più rilevante per l'argomento di questo testo, perché esso rappresenta una tradizione che ha sviluppato una critica della struttura grammaticale del soggetto nominativo dall'interno della propria tradizione — sia pure attraverso mezzi concettuali e religiosi molto diversi da quelli dell'analisi linguistica comparata.

La dottrina buddhista dell'**anatman** (in pali: **anatta**) — il non-sé — è la tesi che non esista un soggetto permanente e sostanziale che sia il «vero» agente delle azioni: il soggetto nominativo che la grammatica delle lingue indoeuropee (e del sanscrito e del pali) presuppone non corrisponde a nessuna entità reale. Ciò che si chiama «sé» è un aggregato impermanente di processi — i cinque *khandha*: forma, sensazione, percezione, formazioni mentali, coscienza — senza un nucleo sostanziale che li unifichi.

La dottrina dell'**anatman** pone immediatamente il problema della responsabilità: se non c'è un sé permanente, chi è responsabile delle proprie azioni? Il karma — la dottrina che le azioni producono conseguenze per il soggetto che le compie — sembra presupporre la persistenza di un soggetto che riceve le conseguenze. La tradizione buddhista ha sviluppato risposte sofisticate a questo problema — la dottrina della continuità della coscienza senza identità sostanziale, la dottrina del flusso causale che connette azioni e conseguenze senza un portatore permanente — ma queste risposte sono state elaborate con fatica e producono tensioni all'interno del sistema, precisamente perché si scontrano con la struttura grammaticale delle lingue in cui la dottrina viene formulata.

Il punto cruciale, dal punto di vista di questo testo, è che il buddhismo ha percepito un problema con la struttura del soggetto nominativo come categoria filosofica — il soggetto grammaticale non corrisponde a una realtà metafisica — senza disporre della risorsa grammaticale alternativa che il georgiano offre. Il georgiano non elimina il soggetto: lo articola in modi diversi a seconda delle dimensioni dell'agentività. Il buddhismo elimina il soggetto come entità sostanziale: esso compie una

mossa più radicale, ma lo fa perché non dispone di una grammatica alternativa che articoli l'agentività in modo non-sostanziale senza eliminare il soggetto.

Se la dottrina del non-sé si fosse sviluppata in una lingua con struttura morfologica analoga al georgiano, la sua formulazione avrebbe potuto essere diversa: non «non vi è un sé permanente», ma «l'agentività è distribuita lungo dimensioni che non convergono in un unico soggetto stabile». Questa formulazione è meno radicale della dottrina buddhista classica, ma potrebbe essere più difendibile filosoficamente — e ha il vantaggio di connettere la critica all'agentività sostanziale con una struttura positiva di attribuzione della responsabilità che il buddhismo ha faticato a sviluppare.

13.10 La traduzione del buddhismo in cinese e le conseguenze grammaticali

La tradizione buddhista chan (zen) cinese è il luogo in cui la grammatica di una lingua alternativa ha effettivamente prodotto trasformazioni nella formulazione della dottrina buddhista che confermano, in modo storicamente concreto, la tesi grammaticale di questo testo.

Il cinese classico — come si è notato nella premessa di questa monografia — non grammaticalizza il tempo verbale in modo morfologicamente obbligatorio nel modo in cui lo fanno le lingue indoeuropee. Esso ha un sistema aspettuale che marca la completezza o incompletezza degli eventi, ma non un sistema temporale che ancora ogni proposizione al presente del parlante come punto di riferimento privilegiato. Questa differenza grammaticale ha prodotto, nella traduzione dei testi buddhisti dal sanscrito e dal pali al cinese classico, trasformazioni che i commentatori hanno spesso attribuito a differenze culturali o religiose, ma che hanno anche una componente grammaticale.

In particolare, la formulazione cinese della dottrina del non-sé è grammaticalmente diversa dalla formulazione indiana. In sanscrito, la proposizione «non vi è un sé» ha la struttura di una negazione dell'esistenza di un soggetto nominativo: essa nega l'entità che la grammatica presuppone. In cinese classico, l'equivalente ha una struttura diversa: non è una negazione di esistenza di un soggetto, ma una proposizione su come i processi si svolgono — «i processi si svolgono senza che vi sia un centro permanente che li coordini». Questa formulazione è

meno paradossale in cinese di quanto non lo sia in sanscrito o in greco, perché la grammatica cinese non presuppone il soggetto nominativo permanente nel modo in cui lo presuppongono le lingue indoeuropee.

Il risultato è che la tradizione zen cinese ha sviluppato la dottrina del non-sé in una direzione più pragmatica e meno metafisicamente paradossale di quella delle tradizioni theravada dell'Asia meridionale: anziché elaborare dottrine complesse sull'impermanenza del soggetto come risposta al paradosso grammaticale, la tradizione zen ha potuto orientarsi verso la pratica — verso l'esperienza diretta dell'azione senza attaccamento al soggetto agente — perché la sua grammatica non le poneva lo stesso problema. Questo è un caso concreto in cui la grammatica ha co-determinato la direzione di sviluppo di una tradizione filosofica — non determinandola, ma modellando il terreno su cui la tradizione si è formata.

VI. Sintesi: il pattern delle tracce grammaticali

13.11 Ciò che i case studies mostrano

I case studies presentati in questa sezione mostrano un pattern coerente che supporta la tesi storico-ricostruttiva di questo testo: nelle tradizioni filosofiche che si sono formate in lingue con struttura binaria dell'agentività, il problema dell'agentività e della responsabilità emerge sistematicamente in forme che riflettono la presupposizione grammaticale del soggetto nominativo unitario. Le distinzioni che la quadripartizione dimensionale proposta in questo testo identifica come filosoficamente rilevanti emergono invece in modo frammentato — in Aristotele, in Anselmo, nella fenomenologia di Merleau-Ponty, nella critica hegeliana al kantismo — senza mai essere sistematizzate in un quadro coerente, precisamente perché la struttura grammaticale delle lingue in cui queste tradizioni si formano non le grammaticalizza come sistema.

Il caso buddhista è il più istruttivo perché mostra il punto di massima pressione: la tradizione buddhista ha percepito il problema con il soggetto nominativo come categoria filosofica, lo ha tematizzato esplicitamente come problema metafisico (il problema dell'anatman), e ha sviluppato soluzioni radicali (la negazione del sé) che hanno il pregio di riconoscere il problema ma il limite di non offrire una struttura alternativa positiva per

pensare l'agentività senza il soggetto nominativo unitario. La risorsa grammaticale del georgiano — che articola l'agentività in modo multidimensionale senza eliminare il soggetto — offre una terza via tra la presupposizione indoeuropea del soggetto nominativo unitario e la negazione buddhista del soggetto.

Questa terza via non è una soluzione definitiva: essa è un programma di ricerca, un orientamento verso una filosofia dell'agentività più adeguata alla complessità effettiva dell'esperienza umana. Ma il fatto che possa essere identificata come terza via — e che la terza via sia grammaticalmente realizzabile in lingue esistenti — è uno degli argomenti più forti per la tesi che la struttura binaria della tradizione dominante non sia l'unica struttura possibile né la più adeguata.

13.12 Una nota sul metodo comparativo

Si deve riconoscere, in conclusione di questa sezione, che il metodo storico-ricostruttivo qui adottato ha limiti che è importante non trascurare. I case studies presentati mostrano corrispondenze strutturali tra grammatica e filosofia, ma non provano relazioni causali: non si può escludere che le stesse strutture filosofiche si sarebbero sviluppate in grammatiche diverse, né si può provare che le strutture filosofiche diverse che si sono sviluppate in grammatiche diverse siano il risultato della grammatica piuttosto che di altri fattori culturali, storici e sociali.

Il valore dei case studies non è dimostrativo ma euristico: essi mostrano dove guardare, identificano pattern che meritano analisi più approfondita, e suggeriscono connessioni che potrebbero essere investigate con metodologie diverse — storiche, psicologiche, antropologiche. La tesi grammaticale di questo testo non si regge sui case studies come su prove: si regge sull'argomento filosofico concettuale del terzo tipo delineato nella prima parte, e usa i case studies come illustrazioni e come fonti di ipotesi.

La storia della filosofia è il laboratorio più ricco a disposizione per chi voglia comprendere come il pensiero umano si sia confrontato con i problemi fondamentali dell'agentività e della responsabilità. Che quella storia mostri pattern coerenti con la tesi grammaticale non la prova, ma la rende plausibile in un senso che nessun argomento puramente concettuale potrebbe fare da solo.

Conclusioni: La Grammatica come Punto di Partenza Invisibile

I. Un bilancio delle tesi

10.1 Cosa è stato mostrato

Questo testo ha sviluppato un argomento in quattro fasi che vale la pena di ricostruire sinteticamente, distinguendo le tesi forti da quelle più esposte a revisione.

La tesi più robusta e meno contestabile è quella storico-ricostruttiva: la tradizione analitica ha sviluppato l'analisi dell'agentività e della responsabilità all'interno di un quadro concettuale che tratta l'agentività come proprietà binaria, e questa struttura binaria è tracciabile — in modo storico-testuale — alla struttura grammaticale del soggetto nominativo nelle lingue in cui quella tradizione si è formata. Questa tesi non richiede alcuna assunzione empirica sulla cognizione e sulla psicologia: richiede solo la lettura attenta dei testi e la comparazione con strutture grammaticali alternative. È la tesi più difendibile e quella che produce le implicazioni filosofiche più immediate.

La tesi filosofico-concettuale — che il sistema morfologico georgiano articoli distinzioni dell'agentività indipendentemente motivate da ragioni filosofiche — è anch'essa robusta, con la precisazione metodologica già enunciata: le quattro dimensioni proposte sono una costruzione filosofica ispirata dalla morfologia georgiana, non una lettura diretta di essa, e potrebbero essere contestate o revisionate sulla base di analisi alternative sia della morfologia che del dominio filosofico.

La tesi applicativa — che la critica strawsoniana e la decomposizione del dibattito compatibilismo/incompatibilismo derivino naturalmente dal quadro multidimensionale — è difendibile, con alcune cautele. La critica a Strawson è la più forte: essa mostra in modo dettagliato dove il sistema strawsoniano produce sottodifferenziazione, e le alternative fenomenologiche proposte sono riconoscibili e motivate. La decomposizione del dibattito sul libero arbitrio è più moderata nei propri risultati rispetto a quanto l'impostazione potrebbe far

sperare: essa non risolve il dibattito, ma lo ristruttura in modo da rendere visibili dove il compatibilismo ha già vinto e dove le questioni rimangono aperte. Questo è un risultato reale, ma meno rivoluzionario di un'eliminazione del dibattito.

10.2 Le obiezioni principali e le risposte

Si possono formulare almeno quattro obiezioni principali all'argomento complessivo, e vale la pena di rispondervi esplicitamente.

Prima obiezione: la quadripartizione è arbitraria. Perché esattamente queste quattro dimensioni e non altre? Fischer e Ravizza distinguono tra il tipo di responsività alle ragioni e il fatto che il meccanismo sia «proprio» dell'agente. Wolf distingue tra la capacità di vedere il bene e la capacità di agire in conformità ad esso. Queste alternative non sono quattro ma due; e potrebbero essere sei o otto. La scelta del georgiano come modello è una scelta tra molte possibili, e non vi è ragione per privilegiare la partizione georgiana rispetto alle alternative.

Risposta: La quadripartizione proposta non è difesa sulla base di essere l'unica corretta, ma sulla base di essere indipendentemente motivata rispetto alle questioni filosofiche che il testo affronta. Essa cattura distinzioni che sono riconoscibili nell'esperienza morale ordinaria e che emergono da contesti filosofici diversi: la distinzione tra originarietà causale e intenzionalità emerge dall'analisi anscombeana e davidsoniana; la distinzione tra intenzionalità e controllo emerge dalla discussione sull'akrasia e sulla coazione; la distinzione tra controllo e responsabilità attribuibile emerge dalla discussione sulla responsabilità vicaria e strutturale. Il georgiano non produce la quadripartizione: la rende visibile come sistema coerente. E l'argomento principale — che il sistema binario lascia indistinte distinzioni rilevanti — non dipende dal numero esatto di dimensioni: dipende dal fatto che ce ne siano più di una.

Seconda obiezione: Strawson può assorbire le distinzioni.

La «qualità della volontà» può essere intesa in modo abbastanza ampio da includere distinzioni tra dimensioni diverse. Le scuse strawsoniane già differenziano tra tipi diversi di attenuazione degli atteggiamenti reattivi. I successori di Strawson — Wallace, Scanlon, Wolf — hanno introdotto distinzioni che si avvicinano alle dimensioni proposte in questo testo. Dunque il sistema strawsoniano è più ricco di quanto si sostenga.

Risposta: Questa obiezione ha una parte corretta — il sistema strawsoniano è più ricco di una caricatura binaria — e una parte che non regge. La parte corretta è che Strawson e i suoi successori hanno introdotto sfumature che non si devono ignorare: la critica di questo testo non è a una caricatura di Strawson. La parte che non regge è che quelle sfumature sono trattate, nel sistema strawsoniano e nei suoi sviluppi, come variazioni di intensità entro una categoria binaria, non come dimensioni distinte con logica propria. La risposta a questa obiezione è mostrarla con casi concreti — e questa sezione ha cercato di farlo: i casi del chirurgo, del genitore, dell'esecutore di ordini, del costruttore di strutture, sono casi che il sistema strawsoniano non riesce ad analizzare con precisione anche nella sua versione più sofisticata.

Terza obiezione: l'argomento è circolare. Si usa il georgiano per motivare la quadripartizione, e si usa la quadripartizione per criticare il sistema strawsoniano. Ma se la quadripartizione fosse motivata indipendentemente, il georgiano sarebbe superfluo; e se il georgiano fosse la motivazione unica, la quadripartizione dipende da una scelta linguistica arbitraria.

Risposta: Il georgiano e la quadripartizione svolgono funzioni argomentative diverse, e questa differenza risponde all'obiezione. Il georgiano svolge una funzione euristica: esso rende visibili distinzioni che altrimenti rimangono implicite, e mostra che un sistema morfologico coerente può articularle. La quadripartizione è poi difesa sulla base di ragioni filosofiche indipendenti: essa non dipende logicamente dal georgiano, ma ne è ispirata. Le due cose non si circondano: il georgiano apre la possibilità della quadripartizione come sistema coerente, le ragioni filosofiche la giustificano come sistema adeguato.

Quarta obiezione: l'argomento storico-ricostruttivo non è dimostrato. Si afferma che la struttura binaria della tradizione analitica è tracciabile alla grammatica delle lingue indoeuropee, ma questa affermazione richiede una storia controfattuale — come sarebbe stata la tradizione se si fosse sviluppata in lingue diverse — che è per definizione impossibile da verificare.

Risposta: Questa obiezione è corretta nel richiamare l'impossibilità della storia controfattuale, ma sopravvaluta ciò che l'argomento storico-ricostruttivo richiede. Esso non afferma che la tradizione sarebbe stata diversa se le lingue fossero state diverse: afferma che la struttura grammaticale delle lingue in cui la tradizione si è formata ha reso certi problemi più naturali da formulare in certi modi, e che questo

si vede nei testi effettivi. Questa è un'affermazione interpretativa, non una storia controfattuale: si regge sulla qualità dell'interpretazione testuale, non su esperimenti mentali impossibili.

10.3 Il programma di ricerca

Le tesi di questo testo, se corrette, aprono un programma di ricerca in diverse direzioni che non è possibile sviluppare qui, ma che vale la pena di indicare.

Sul fronte linguistico, è necessario sviluppare l'analisi del sistema georgiano con il rigore tecnico che una critica cauoascolgica richiede, e estendere la comparazione tipologica in modo sistematico. Lingue come l'avar, il tibetano, il tzeltal, il quechua andrebbero analizzate rispetto alla grammaticalizzazione dell'agentività con la stessa cura dedicata al georgiano, per verificare se le distinzioni identificate siano tipologicamente convergenti.

Sul fronte della psicologia morale, la tassonomia differenziata degli atteggiamenti reattivi abbozzata in questo testo richiede sviluppo fenomenologico e verifica empirica. Si potrebbero disegnare studi cross-culturali per verificare se parlanti di lingue che grammaticalizzano le dimensioni dell'agentività in modo diverso abbiano intuizioni morali strutturalmente diverse rispetto ai casi multidimensionali identificati.

Sul fronte della filosofia del diritto, la struttura multidimensionale della responsabilità potrebbe illuminare problemi pratici della teoria della responsabilità penale, civile e amministrativa che i sistemi giuridici vigenti trattano in modo ad hoc perché mancano del quadro concettuale adeguato.

10.4 Una nota finale sulla grammatica e il pensiero

Questo testo si è sviluppato sotto la guida di un'idea che vale la pena di esplicitare in forma conclusiva. La grammatica non è il destino del pensiero: nessuna lingua impedisce ai propri parlanti di pensare ciò che la struttura grammaticale non rende automatico. Ma la grammatica determina il punto di partenza del pensiero — le categorie che sono disponibili senza sforzo, le distinzioni che si fanno automaticamente, i problemi che si vedono prima di quelli che richiedono analisi deliberata. E il punto di partenza determina, nel lungo periodo e attraverso le

generazioni di una tradizione intellettuale, quali percorsi sono frequentati e quali rimangono raramente battuti.

La filosofia della responsabilità si è sviluppata su un terreno grammaticale che rendeva facile la struttura binaria dell'agentività e difficile la sua articolazione multidimensionale. Questo non è un difetto dei filosofi che vi hanno contribuito: è una struttura del medium in cui hanno lavorato. Renderla visibile non è un atto di critica verso quella tradizione: è un atto di comprensione della sua forma, che permette di vedere dove i percorsi raramente battuti potrebbero condurre.

I percorsi raramente battuti, in questo caso, conducono verso una teoria della responsabilità più adeguata alla complessità effettiva dell'agentività umana — una complessità che le grammatiche di certe lingue hanno avuto il merito, o la fortuna, di articolare con chiarezza morfologica da prima che la filosofia riuscisse ad articolarla con chiarezza concettuale.